International Journal of Approximate Reasoning 53 (2012) 636–659

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

A skin detection approach based on the Dempster–Shafer theory of evidence

Mohammad Shoyaib [a], M. Abdullah-Al-Wadud [b], Oksam Chae [a,*]

[a] Department of Computer Engineering, Kyung Hee University, 1 Seochun-dong, Kihung-gu, Yongin-si, Kyunggi-do 446-701, Republic of Korea
[b] Department of Industrial and Management Engineering, Hankuk University of Foreign Studies, 89 Wangsan, Mohyun, Cheoin, Yongin, Gyonggi 449-791, Republic of Korea

ARTICLE INFO

Article history:
Received 9 August 2010
Received in revised form 13 January 2012
Accepted 16 January 2012
Available online 28 January 2012

Keywords:
Dempster–Shafer theory of evidence
Image processing
Pattern recognition
Skin detection

ABSTRACT

Skin detection is an important step for a wide range of research related to computer vision and image processing and several methods have already been proposed to solve this problem. However, most of these methods suffer from accuracy and reliability problems when they are applied to a variety of images obtained under different conditions. Performance degrades further when fewer training data are available. Besides these issues, some methods require long training times and a significant amount of parameter tuning. Furthermore, most state-of-the-art methods incorporate one or more thresholds, and it is difficult to determine accurate threshold settings to obtain desirable performance. These problems arise mostly because the available training data for skin detection are imprecise and incomplete, which leads to uncertainty in classification. This requires a robust fusion framework to combine available information sources with some degree of certainty. This paper addresses these issues by proposing a fusion-based method termed Dempster–Shafer-based Skin Detection (DSSD). This method uses six prominent skin detection criteria as sources of information (SoI), quantifies their reliabilities (confidences), and then combines their confidences based on the Dempster–Shafer Theory (DST) of evidence. We use the DST as it offers a powerful and flexible framework for representing and handling uncertainties in available information and thus helps to overcome the limitations of the current state-of-the-art methods. We have verified this method on a large dataset containing a variety of images, and achieved a 90.17% correct detection rate (CDR). We also demonstrate how DSSD can be used when very little training data are available, achieving a CDR as high as 87.47% while the best result achieved by a Bayesian classifier is only 68.81% on the same dataset. Finally, a generalized DSSD (GDSSD) is proposed achieving 91.12% CDR.

1. Introduction

Skin detection is performed as a preliminary step in most human-related image processing applications. The detected skin regions are then usually further processed based on the application focus such as face detection, gesture recognition, web content filtering (e.g., pornographic filters) and video surveillance applications. The performance of these applications is significantly affected by the accuracy of the skin detection step.

It is well agreed that correctly modeled and represented skin clusters in a color space can serve as an effective tool for skin detection [1]. However, various factors make skin detection challenging in real life applications such as variations in illumination, ethnicity, shadow, hairstyle, background, makeup and camera characteristics.

* Corresponding author.
E-mail addresses: shoyaib@khu.ac.kr (M. Shoyaib), wadud@hufs.ac.kr (M. Abdullah-Al-Wadud), oschae@khu.ac.kr (O. Chae).

0888-613X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2012.01.003

Proper handling of these variations requires an accurate model for skin and non-skin, which is not achievable in practice. This is because the available training data are imprecise and incomplete, which leads to uncertainty in decisions [82]. Example 1 presents a scenario to explain the skin classification problem.

Example 1. Consider a classifier that can classify a given color as either skin or non-skin based on its value in only the R channel of the RGB color model. The training data will be precise if the set of all possible R-values of skin samples and that of non-skin samples are disjoint. However, several color values are found in both skin and non-skin training samples. Based on these data, this classifier can at best predict whether a given pixel color is skin or not, with some uncertainty. This means that the classifier is not fully reliable. Further, it is usually possible to obtain a sufficient variety of samples to represent the array of natural skin tones, but the set of non-skin data is huge, and it is impossible to collect all possible samples. Suppose that for $R = 160$ the training dataset only includes skin samples. The fact that we do not have all possible non-skin samples in the training set means that it is possibility that some non-skin samples will have $R = 160$. This lack of sample can lead to incompleteness in the training data.

Several methods have been proposed [2,38,42] to model skin and non-skin colors. However, these methods may fail to generate optimal results since they may lack proper uncertainty management. To reduce the uncertainty these methods usually use different types of parameters and thresholds, which in turn introduce additional problems such as a long training/testing time and/or difficult parameter and threshold tuning.

These methods generally detect skin using color information that may not always provide completely decisive and reliable information. In this scenario, combining information from different sources may yield a more reliable decision. Such information fusion has become highly popular for many applications related to image processing and computer vision [62] because it improves the quality of a decision by decreasing uncertainty and imprecision and increasing the amount of global information [65].

In general, all the current state-of-the-art methods exhibit poor performance when trained with a small amount of training data. For example, as presented in [2], the skin detection accuracy of the Bayesian-based method drops from 90% to approximately 77% (for the same false detection rate) when it is trained with 1% of the full training data. However, the ability to train with small amounts of data has some practical importance. For instance, any surveillance system installed on a client site may require training with a few local images, and human detection and tracking in abnormal light (such as in a specific color of light or under specific illumination) may require training using images from the on-site lighting environment for better performance. In such situations, it is necessary to train the system with the few available images within a short period of time to ensure acceptable performance.

To overcome the aforementioned limitations, this paper introduces a novel fusion-based method named Dempster–Shafer-based Skin Detection (DSSD) that focuses on the following issues for processing available information meaningfully to attain acceptable detections:

- Selection of good sources of information (SoIs).
- Determination of mass values (confidences) of the selected SoIs.
- Combining the mass values to obtain the final decision.

To handle these issues, the proposed DSSD identifies a few well-suited skin detection criteria (SoIs) by analyzing the color distributions of skin and non-skin pixels. We also propose a way to quantify their confidences in classifications. These individual confidences are then combined using Dempster–Shafer Theory (DST) [5] to obtain the final decision. We choose DST for this skin detection method because it offers a suitable framework to combine the decisions of different SoIs to handle uncertainties [66].

The experimental results show that the proposed DSSD outperforms the currently available techniques (higher accuracies, lower training time). Moreover, DSSD also shows acceptable accuracy when trained with a small amount of data. Finally, we extend DSSD to a generalized DSSD (GDSSD), which shows better performance than DSSD and requires neither threshold nor parameter selection.

The rest of the paper is organized as follows: Section 2 presents the currently available approaches for skin detection. Section 3 describes the Dempster–Shafer Theory of Evidence. Some preliminary considerations such as which color space is used and how SoIs are selected are then described in Section 4. The DSSD and GDSSD are presented in Section 5. Section 6 analyzes the performance of these approaches with that of other well-known algorithms, and Section 7 presents the conclusions of this study.

2. Related work

Skin detection can be considered a binary classification problem, i.e., whether a given pixel color represents skin or not. The existing skin detection methods can be categorized into two broad categories based on whether any training session(s) is required or not: statistics-based methods and explicit threshold based methods (ETM), respectively. The ETMs use some predefined decision criteria, whose outcomes are combined using the logical AND operation to represent skin clusters in color space. The statistics-based methods can be further categorized into three groups: parametric, semi-parametric and

non-parametric [38]. Parametric skin color models use some specific functional forms (such as Gaussian) and thus adjust the required parameters (such as mean and covariance) to approximate the models based on the training datasets. Semi-parametric methods do not have such specific functional forms. They use the general form of adaptive parameters/weights (such as a neural network based method) that must be tuned according to the training data to represent the models [38]. Non-parametric methods (such as the histogram thresholding method used in [2]) do not use any particular functional form or weight. There are also some fusion-based methods in which the results of different skin detection methods are combined to make the final decision.

Table 1 presents the performances, in terms of CDR and FDR as defined in Section 6.2, of some available skin detection methods reported by various authors. Note that these results are obtained from different training and test datasets from same/different databases, and thus different authors report different results, even when using the same method. Therefore, this table provides an overview rather than a fair comparison of the performances of the various methods. The following subsections also present a short overview of some of the popular methods in the above-mentioned categories such as statistics-based methods, explicit threshold-based methods and fusion-based methods along with their pros and cons, and then we briefly describe our proposed contributions.

2.1. Statistics-based methods

This section briefly describes some parametric, non-parametric and semi-parametric statistical methods.

2.1.1. Parametric methods

Parametric methods include single Gaussian [23,26–28], Mixture-of-Gaussian (MoG) [11,15,24,29,30,50] and elliptical boundary models [25]. These models require few training samples to obtain well-generalized models, and their memory requirements are also low. Expectation maximization (EM) is usually used to approximate the parameters (for instance, the mean, covariance matrix and weight) of the models that fit the training data [23,29,50].

The methods that use a Gaussian model consider skin to have a color distribution that is clustered in a chromatic color space and can be represented by a Gaussian law. However, the skin cluster has a complex boundary in the color space that these models cannot accurately represent. Mixture-of-Gaussian (MoG) models consider a set of Gaussian components that when combined may be able to handle different conditions, especially different luminance levels [2]. Such combinations give better results in poor or strong lighting conditions, but a change in illumination is more complex in real life. Moreover, these methods require expensive training and a long classification period [2]. Lee et al. [25] propose an elliptical boundary model

Table 1
Performance of existing skin detection methods.

Category	Method	Color space	Database	CDR	FDR
Non- parametric	Bayesian [2]	RGB	Compaq [2]	90.00	14.20
	Bayesian [89]	RGB	Compaq	93.40	19.80
	Bayesian [48]	RGB	ECU [48]	≈83.00	10.00
	Adaptive Bayesian [52]	RGB	Compaq	69.00	5.00
	Bayesian [52]	RGB	Compaq	74.50	5.00
	Bayesian [78]	RGB	N/A*	96.65	N/A
Parametric	MoG [2]	RGB	Compaq	90.00	15.50
	Elliptical model [25]	Xyz	Compaq	90.00	20.90
	MoG[25]	YIQ	Compaq	90.00	30.00
	MoG [48]	YCbCr	ECU	≈81.00	10.00
	MoG [77]	HSV	IBTD[2]	92.00	37.80
	MoG [90]	rgb	N/A	87.00	30.00
	MoG [78]	YCbCr	N/A	99.10	N/A
Semi-parametric	SOM [42]	TSL	Compaq	78.00	32.00
	MLP [48]	RGB	ECU	≈82.00	10.00
	SVM + MoG [77]	HSV	IBTD	92.00	30.00
ETM	Threshold [48]	CbCr	ECU	≈93.00	29.09
	Threshold [4]	HSI	Compaq	93.00	≈24.00
	Threshold [4]	RGB		89.00	≈21.00
	Threshold [4]	YCbCr		90.00	≈34.00
	Threshold [4]	HSV		46.00	≈7.00
	Threshold [4]	rgb		42.00	≈11.00
	Threshold [78]	RGB	N/A	100.00	N/A
Fusion	Sum [4]	RGB, rgb, YCbCr, HSV, HSI	Compaq	≈98.00	≈42.00
	Majority [4]			≈82.00	≈18.00
	Product [4]			≈32.00	≈4.00
	SCNS [4]			≈92.00	≈22.00

* In the table, N/A stands for "not available", and ≈ means approximated values from plots.

to reduce the complexity and keep the detection accuracy comparable to MoG. This model also requires a long classification time. Bin Li et al. [51] propose an iterative process for skin region segmentation. This method first detects some skin pixels from a test image based on a generic MoG model and then keeps updating the model based on the detected pixels. This process shows good performance for many real life images. However, the system is designed under the assumption that the number of training images is sufficiently large so that the trained model contains a posterior probability similar to the test images, and hence that the model can initially locate the skin cluster in different images in the first step, which is not always true. Furthermore, this model requires many parameters to be tuned, making the detection process slower.

2.1.2. Non-parametric methods

Non-parametric skin detection [2,9,32–36,42] methods estimate the skin color distribution from the histogram of the training data. The Bayesian decision rule for minimum cost is an effective technique in the statistical pattern classification area. Jones and Rehg [2] use this rule to classify image pixels as skin or non-skin. This method established a threshold for the ratio of the likelihood that a pixel is skin to the likelihood that it is non-skin to decide whether a given pixel is skin or not. Bayesian based methods (e.g., [2]) are most frequently used for their computational simplicity and acceptable accuracies (see Table 1).

Parametric methods make use of approximated parameters that define the approximate boundaries of the skin clusters. In contrast, a Bayesian classifier does not make any approximation. It represents the models (histograms) as they are trained, and applies a threshold to the values to produce a complex boundary of skin clusters. The advantage of using a Bayesian classifier is its higher detection rate compared with the other classifiers. This is due to the fact that other classifiers cannot represent the complex skin cluster boundary in the color space. A Bayesian classifier will work acceptably if the data for two classes can be modeled accurately. However, an accurate statistical model theoretically requires an infinite amount of training data encompassing all non-skin colors in nature, and the performance of the classifier surely suffers in the absence of an exact model of all possible non-skin colors. Therefore, no perfect statistical model exists yet. Moreover, such methods require substantial memory to store two models — one for the distribution of colors in skin samples and another for the distribution of non-skin colors.

2.1.3. Semi-parametric methods

The artificial neural network (ANN)-based skin detection methods mainly fall in the semi-parametric category. In [42], Brown et al. propose an ANN model using a self organizing map (SOM). They use two separate SOMs for skin and non-skin distributions and report better performance than MoGs. However, the performance of this model is worse than that of the Bayesian classifier presented in [2]. Detection methods based on a multilayer perceptron (MLP) [38–41] can also represent a skin model with a complex boundary. However, these methods may not be able to cope with missing information such as non-skin samples absent from the training set [1]. Moreover, the training time for MLP models is much longer than that for the histogram-based approach in the presence of a large amount of training data [48]. Recently, a deep multi-layer neural network (DMLNN) [76] model was proposed that has become popular for various classification problems. Unlike a traditional ANN, this model contains many levels to represent highly nonlinear and highly varying functions. However, like Bayesian-based methods, most of the existing ANN-based skin detection methods also suffer from inadequate amounts of non-skin data. Furthermore, they also require very long training and classification periods and exhibit slow performance.

The support vector machine (SVM) has become popular for its performance in the machine learning area. The SVM method is expected to perform the best when only the features relevant to the focus of classification are available. Too many extraneous inputs may harm the results obtained on real data [85]. SVM for skin detection may also suffer from these shortcomings because it is difficult to determine which features are relevant as we do not have all non-skin samples and many regions overlap between skin and non-skin clusters. Furthermore, SVM can take a very long time for training and classification. Zhu et al. [77] apply SVM to classify the MoG parameters of skin and non-skin models, not the pixel colors. This approach minimizes the amount of training data required and yields relevant information from MoG parameters. These authors first use a generic MoG model to identify most of the skin-like pixels in an image, including some non-skin pixels as well. These skin-like pixels are then refined by the MoG model with two kernels: one for the true skin model and another for the false skin model. Finally, they use SVM to classify the MoG parameters of these two models. The main advantage of this method is that it is adaptive to encompass different lighting conditions. However, in this case, the SVM decision is based on the outputs of the MoG. Therefore, this method has similar problems as the standard MoG-based skin detection methods mentioned in Section 2.1.1 such as computational expense in training and run time.

2.2. Explicit threshold-based methods

Among the methodologies developed to date, the ETMs are the simplest for classifying skin and non-skin pixels [1,4,18]. These methods explicitly define the boundaries of the skin cluster in certain color spaces using a set of fixed thresholds [4,6–10,13,20]. The benefit of such methods is that they can be used right away without requiring any training phase. In these cases, few criteria are defined and each criterion (boundary) yields many false positive pixels when applied individually as the thresholds are set to be large enough to capture all variations in color. Such methods combine the outcomes of the criteria by applying logical AND operations. This helps to reduce false positive results but also yields a large decrease in true

positives. Moreover, all the criteria are given the same weight. If any of the criteria misclassifies a skin pixel then the error is simply propagated to the final output. Moreover, since these approaches are guided by some constant values, they may lack the flexibility required to work under a variety of conditions.

2.3. Fusion-based methods

Recently, fusion-based methods have gained a great deal of attention in computer vision and image processing. This is because uncertainty is an inherent characteristic of the available information extracted from images, and fusion helps to reduce this uncertainty [65]. Dargham et al. [49] propose to use the AND operator for fusing information. One of his proposed methods fuses two chrominance components from the same color space, and the other fuses the output of two different skin detection methods. However, such fusion methods show poor performance with real data. Gasparini et al. [4] describe four different fusion rules, namely the sum rule, the product rule, the majority vote and the SCNS (skin corrected by a non-skin) rule to combine the outcomes of some skin detection methods. They also propose a white balance algorithm as a preprocessing step to enhance the detection rate. However, most of the fusion rules perform poorly (see Table 1) and are not robust enough to achieve state-of-the-art performance. These authors also agree that their proposals do not exhibit better performance than histogram-based algorithms.

Sun [52] describes a skin detection method that combines global (histogram-based learning) and local information (collected using k-means clustering) to work in different conditions. However, this method requires many parameters to be tuned manually. Moreover, like other Bayesian skin detection methods, it also requires a large dataset for training.

Fusion-based methods try to combine information from different sources to arrive at more reliable decisions. However, the fusion processes described above are not able to deal with uncertainty. To combine information from multiple sources and improve classification accuracies, we require a more robust and flexible fusion framework.

2.4. Proposed contributions

Most of the currently available statistics-based methods (and all ETMs) make the decision to identify skin depending on one or a few thresholds. Fixed thresholds are used in ETM methods. On the other hand, although Mixture of Gaussian (MoG) and Bayesian-based classifiers[1] do not usually use thresholds to generate the models, these methods are highly dependent on the threshold values that are used to make final decisions. Lowering (increasing) the values of the thresholds increases (decreases) the true positives, and also increases (decreases) the false positives. Therefore, setting a general threshold to get very good performance under all conditions may not be achievable in these methods. It is generally recommended to set the skin detection thresholds according to the application. In the case of MLP and other ANN-based methods, the major drawback is the lack of proper uncertainty management to handle incomplete data, and for SVM, it is a challenge to find an appropriate amount of relevant samples.

From the above discussion, it is clear that skin detection could benefit from an improved and reliable approach that can achieve high accuracies even with a very small amount of training data and short training/testing times. Such a method will not only enhance skin detection, but will also accelerate the progress of related research.

The Dempster–Shafer theory (DST) of evidence is a flexible fusion theory that is capable of handling the imprecision and incompleteness present in data and the uncertainties in different information sources [67]. The success of the DST in pattern recognition and classification is addressed by a number of researchers [66,68,69,74] and has been applied in several fields such as moving objects [59], people [60] and face [70] tracking, and facial expression recognition [61]. Some well-known proposals have also described DST-based solutions for classification problems. Some of these proposals follow a case-based approach (e.g., [66,71,72]) while others follow a model-based approach (e.g., [80,81]). The relationship between these two methods is described in [73].

Inspired by the success of DST, we elect to use this method for skin detection. However, one of the major challenges is to find a way to calculate the confidences of different SoIs from the available data. We first focus on the methodologies available in the literature [66,71]. In [71], the reliabilities (mass values) of the SoIs are calculated from the neighboring information according to a distance metric. The performance of this method is quite good when the number of patterns under consideration is small. However, in skin detection, there are $(256 \times 256 \times 256)$ different colors in the RGB color space, so both training and testing require a very long time. Denoeux [66] addresses this problem by using some prototypes instead of all samples. The prototypes can be found using a standard clustering method such as a c-means algorithm. Mass values are calculated based on a distance measure obtained from these prototypes. This process is implemented using a multilayer neural network consisting of one input layer, two hidden layers and one output layer. The weight vector, the receptive field, and the class membership of every prototype are computed by minimizing the mean squared difference between the classifier outputs and the target values. We have applied this method for skin detection (denoted as NNDST) and found reasonable results (demonstrated in Section 6.3). However, this method is still expensive to train for large skin detection training datasets in terms of both computation and memory requirements. The method also requires a long time for classification, especially due to the many distance calculations carried out for each pixel.

[1] Brief descriptions of the Bayesian and MoG-based methods are included in Appendix A.

This paper aims to achieve several goals such as low training time and computational complexity, high accuracy with no threshold and a limited number of parameters requiring tuning, and effective uncertainty management. To meet these requirements, we propose Dempster–Shafer based Skin Detection (DSSD), a fusion-based method that performs a meaningful combination of different sources of information (SoIs) [53]. Thus, even if some SoIs fail to produce an accurate decision, other SoIs can correct the final decision. We also extend DSSD and propose a generalized DSSD (GDSSD), which exhibits excellent performance and requires no threshold specifications. The major contributions of this work are: (1) the proposed mass calculation method (and the determination of boundaries of SoIs in DSSD) requires very little computation, and (2) our methods have the ability to work well with a small amount of training data.

3. Dempster–Shafer theory of evidence

The Dempster–Shafer theory of evidence, which was first stated by Dempster in 1960 and later extended by Shafer [5], is capable of representing uncertainty as well as ignorance in statistical measurements. The method can decrease the amount of uncertainty in information by applying a combination rule to combine the confidences of different information sources, resulting in a more precise definition of the hypotheses.

The DST uses a *frame of discernment*, which is defined as a set of mutually exclusive and collectively exhaustive hypotheses denoted by $\boldsymbol{\Theta}$. Let us suppose $\Theta = \{h_1, \ldots, h_n\}$. The power set of all possible subsets of $\boldsymbol{\Theta}$, including itself and the empty set $\emptyset$, is 2^{Θ}, i.e., $2^{\Theta} = \{\emptyset, \{h_1\}, \ldots, \{h_n\}, \{h_1 \cup h_2\}, \ldots, \{h_1 \cup h_n\}, \{h_2 \cup h_3\}, \ldots, \{h_2 \cup h_n\}, \ldots, \{h_{n-1} \cup h_n\}, \{h_1 \cup h_2 \cup h_3\}, \ldots, \Theta\}$. Usually, some SoIs are selected that are capable of providing distinguishable information for some subsets (S) of $\boldsymbol{\Theta}$, i.e., $S \in 2^{\Theta}$. A mass function (alternatively known as a basic belief or basic probability assignment) m: $2^{\Theta} \to [0,1]$ is a function satisfying

$$\left.\begin{array}{l} m(\emptyset) = 0 \\ m(S) \geqslant 0, \ \forall_{S \subseteq \Theta} \\ \sum\limits_{S \subseteq \Theta} m(S) = 1 \end{array}\right]. \tag{1}$$

Here, $m(S)$ represents the belief provided by an SoI in favor of S exactly and fully. This parameter reflects how strongly the SoI supports S. There is no belief for the empty set $\emptyset$ and all assigned mass values sum to unity. The mass value assigned to, Θ i.e., $m(\theta)$, is called the degree of ignorance and the subsets S *of* Θ of with non-zero mass values are called the focal elements.

Eq. (1) makes this theory different from the probabilistic approach. While probabilistic approaches can only handle singleton focal elements, the DST can handle both compound sets and singletons [43].

Belief (*bel*) and plausibility (*pl*) are two other common evidential measures that are derived from the mass function as given in Eqs. (2) and (3), respectively,

$$\left.\begin{array}{l} bel(\emptyset) = 0 \\ bel(S) = \sum\limits_{T \subseteq S, T \neq \emptyset} m(T) \end{array}\right], \tag{2}$$

$$\left.\begin{array}{l} pl(\emptyset) = 0 \\ pl(S) = \sum\limits_{S \cap T \neq \emptyset} m(T) \end{array}\right], \tag{3}$$

where S and T are subsets of Θ.

bel(S) and *pl*(S) represent the exact and possible support to S, respectively. The interval [*bel*(S), *pl*(S)] can be interpreted as the upper and lower bound of the probability [5,71].

If there are different independent SoIs, DST provides a way to combine the mass values assigned by those SoIs. Dempster's rule of combination fuses the mass functions m_i obtained from n SoIs according to Eq. (4):

$$\left.\begin{array}{l} m(\phi) = 0, \\ m(S) = \frac{\sum_{S_1 \cap \ldots \cap S_n = S} \prod_{i=1}^{n} m_i(S_i)}{1-K} \end{array}\right], \tag{4}$$

where K represents the degree of conflict given by

$$K = \sum_{S_1 \cap \ldots \cap S_n = \emptyset} \prod_{i=1}^{n} m_i(S_i).$$

There are several ways of making the final decision using the DST framework. For instance, the decision can be made by choosing the hypothesis with the maximum mass, belief, or plausibility [21,54,64] or by using a pignistic probability distribution [3,58,63,66].

This paper considers the two class problem of whether a given color represents skin or non-skin. It can be easily shown that all these alternatives will lead to the same output in such cases. Hence, in our work, we make the final decision by selecting the hypothesis that produces the maximum aggregation of the mass values (according to Eq. (4)) using

$$X = \text{argmax}_{A \subseteq \Theta}(m(A)). \tag{5}$$

3.1. DST and skin detection

One of the main reasons for choosing DST for the application of skin detection is that it can deal with ignorance and missing information, which is important as we do not have all possible non-skin samples. In DST, mass values are assigned only to those subsets of the environment for which enough information is available. The remaining mass is kept as undetermined, assigned to Θ comprising all the hypotheses, and used later. This makes DST capable of handling uncertainty much better than the classical probabilistic theory, where the unavailable probability is assigned to the complement of the known set. Example 2 illustrates one such example.

Example 2. Suppose that an expert believes with 70% certainty that a given pixel is skin. Considering this as a probability gives $P(\text{skin}) = 0.7$, and according to classical probability this implies $P(\text{non-skin}) = 0.3$. For the sake of simplicity, let us assume that this 70% certainty also represents the mass value $m(\{\text{skin}\}) = 0.7$. As we do not have any further information about the remaining 30%, DST will assign it to the hypothesis including both skin and non-skin, i.e., $m(\{\text{skin,non-skin}\}) = 0.3$; not to non-skin only.

Classical probabilities are only assigned to elementary hypotheses but DST can consider all possible combinations of hypotheses, making it better able to handle uncertainties than the Bayesian classifier, which is based on classical probability [55]. Again, if $m(\boldsymbol{\Theta}) = 0$ in a two class problem, then both techniques consider only elementary hypotheses. From this, we can easily understand that Bayesian probability theory is a special case of DST [56]. Furthermore, one of the computational advantages of the Dempster–Shafer framework is that priors and conditionals do not need to be specified, unlike Bayesian classifiers. Any information contained in the missing priors and conditionals is not used in the Dempster–Shafer framework unless it can be obtained indirectly. Thus, Dempster–Shafer theory allows a degree of ignorance to be specified in this situation rather than forcing the user to supply prior probabilities. Example 3 shows the superiority of the DST over classical probability in handling ignorance.

Example 3. Suppose that we are concerned with classifying among two classes, W_1 and W_2. An expert x believes with 60% certainty that a given sample falls into W_1. x also believes with 40% certainty that this sample falls into W_2. Another expert y believes with 30% that the given sample falls into W_1. Following an approach similar to Example 2 yields Table 2.

In the case of expert x, both the classical probability and the DST handle the available information similarly. However, for y, there is a clear distinction. At this point the classical probabilistic method assigns the remaining unavailable information $(1 - P(W_1|y))$ to $P(W_2|y)$ while the DST keeps it undistributed.

Thus the posteriori probability would be

$$P(W_1|x,y) = \frac{P(x,y|W_1)P(W_1)}{P(x,y)}.$$

When the experts make independent observations, we can write

$$P(W_1|x,y) = \frac{P(x|W_1)P(y|W_1)P(W_1)}{P(x,y)}.$$

This leads to

$$P(W_1|x,y) = \frac{P(W_1|x)P(W_1|y)}{P(W_1)} = \frac{0.18}{P(W_1)}.$$

Similarly, we get

$$P(W_2|x,y) = \frac{0.28}{P(W_2)}.$$

At this point, the Bayesian-based approach requires complete knowledge about the priors. In the absence of this information the priors are usually considered uniform. Following this approach in our example, the Bayesian-based approach makes the decision in the favor of W_2 even though expert y has provided no information for W_2.

Following Eq. (4), on the other hand, the DST gives $m(W_1) = 0.68$ and $m(W_2) = 0.32$. Thus, the final decision is made in favor of W_1, which fits logically with the two experts' opinions.

Table 2
An example of representing mass values and probabilities.

Expert	Classical probability		Dempster–Shafer theory		
			$m(W_1)$	$m(W_2)$	$m(\boldsymbol{\Theta})$
x	$P(W_1\|x) = 0.6$	$P(W_2\|x) = 0.4$	0.6	0.4	0.0
y	$P(W_1\|y) = 0.3$	$P(W_2\|y) = \mathbf{0.7}$	0.3	**0.0**	**0.7**

The bold numbers show the difference between the assignments done by "Classical probability" and "Dempster-Shafer theory".

Considering the aforementioned discussion and the data we have for skin detection, DST is a good choice for skin detection. Another reason for choosing DST is that there is a large overlap between the distributions of skin and non-skin colors (more details of these distributions are explained in Section 4.2) if plotted with respect to a color channel. Such overlaps are one of the main reasons behind the uncertainties. As the DST framework allows such uncertainties to be handled by the outcomes of the available SoIs, we focus on applying DST to skin detection. The DST approach to handling uncertainties enables DST to function well in reasoning tasks, such as (1) representing incomplete knowledge, (2) belief-updating and (3) evidence pooling.

4. Preliminaries

This section first discusses the color space used in this paper by briefly reviewing color spaces and presenting the reasons for our color space choice, and then describes the procedure used to select SoIs.

4.1. Color space selection

Different color spaces have been proposed to minimize the effects of changes in lighting conditions to make the skin detection process less sensitive to illumination. Some approaches [13,15–19] remove the luminance component and use a 2D representation. However, significant changes in illumination yield significant changes in skin color, even for the same person [12]. In such cases, some color information is also lost while separating luminance, which in turn may degrade the performance. Hence, the computational cost for the transformation from RGB to another color space may not be fruitful. Several renowned researchers reached similar conclusions regarding color space transformation. Fu et al. [57] and Shin et al. [37] have experimented on the suitability of different color spaces for skin detection. Based on their works, Kakumanu et al. [1] conclude that color space transformation does not affect non-parametric models but does affect parametric methods. Experimenting on different color spaces, Phung et al. [48] also recommend using RGB in non-parametric skin classifications.

4.1.1. Color space for SoIs

We have adopted the RGB color space for our work for several reasons. First, the proposed methods (DSSD and GDSSD) are non-parametric. Hence, according to the recommendations of other researchers, we use the RGB color space that ensures simplicity, effectiveness and speed (as it does not require color space transformation). Second, it is reasonable to use the color space that preserves the most information. As the RGB color space retains all color information, we choose to use it for our skin detection. Again, a large amount of training data could encompass a good range of illumination variation [2]. Further, a DST-based framework also helps the system to cope with missing information. In the case of large changes in illumination, however, a preprocessing step also helps to reduce the effect of illumination variation as described later in the paper.

4.2. Selection of the sources of information (SoI)

The colors of skin pixels occupy only a certain area in the RGB color space[2], which is shown in Fig. 1. The primary goal of skin detection techniques is to define the irregular shape of the skin cluster, including as few non-skin colors as possible. In our work, we look for criteria that can distinguish skin from non-skin colors. The selected criteria can be considered as SoIs in DST.

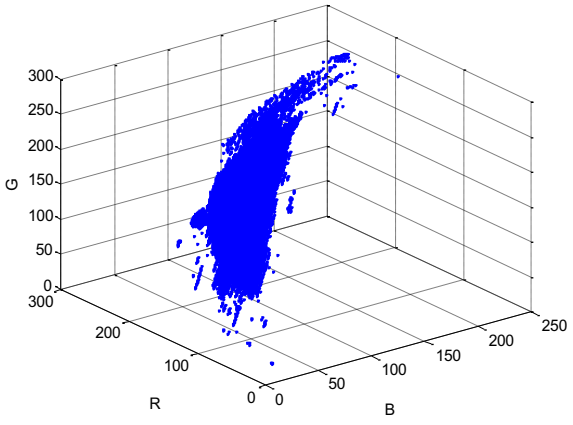


Fig. 1. Skin cluster in RGB color space.

[2] To ensure that the color distributions presented in different figures in this paper are representative, we use a large collection of skin and non-skin images in a huge variety of conditions and normalize the distributions to the total number of skin and non-skin pixels (see Section 6.1 for details).

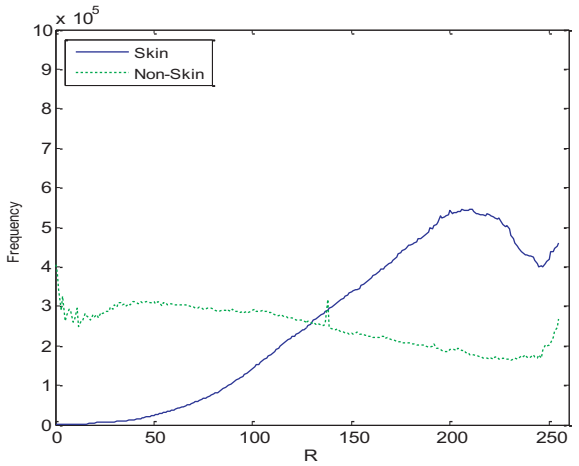
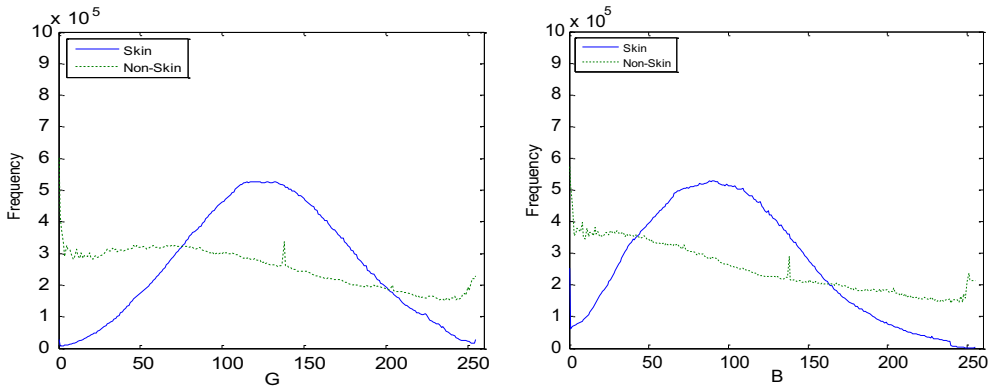


Fig. 2. Distribution of skin and non-skin clusters in R space.

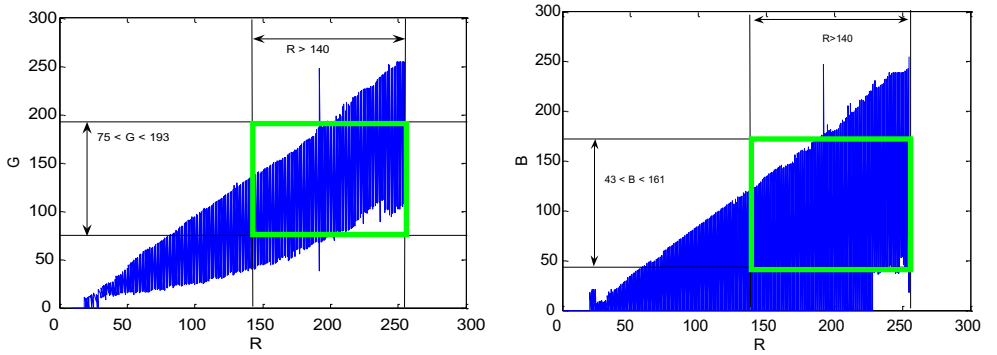


(a) Distribution in G space (b) Distribution in B space

Fig. 3. Distribution of skin and non-skin clusters in G and B space.

The selection of SoIs requires analysis of the color distribution of skin and non-skin pixels. We start from the classification ability of each of the R, G and B channels. The distributions of skin and non-skin colors in R channels are plotted in Fig. 2.

Fig. 2 reveals that a critical value exists of $R = 140$, above and below which the distribution is dominated by skin and non-skin samples, respectively. Hence, we establish the criterion $R > 140$, which indicates a greater chance that the color represents skin when the criterion is satisfied. Similarly, observing the plots in Fig. 3, we select two other criteria as $75 < G < 193$ and $43 < B < 161$, which also demonstrate some capability for distinguishing skin and non-skin pixels.



(a) Distribution in RG space (b) Distribution in RB space

Fig. 4. Distribution of skin colors in RG and RB planes. The rectangular marks show the areas defined by combining the criteria in different color channels.

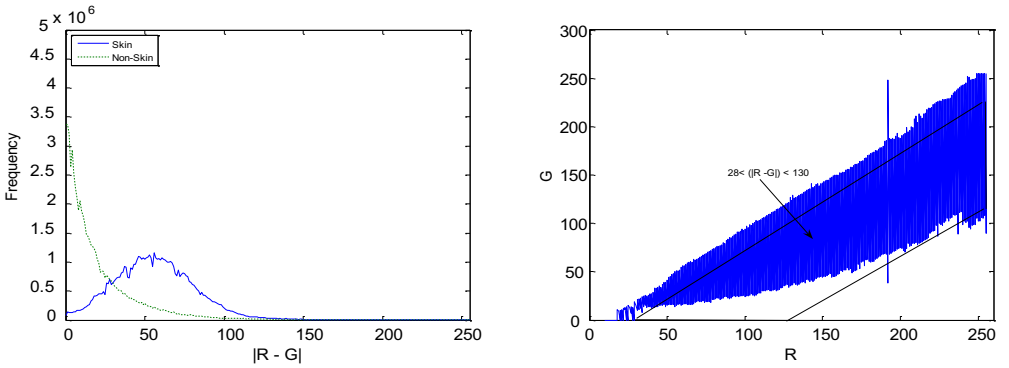
Fig. 4 shows two scenarios that enable the construction of decision boundaries by the joint contribution of R and G or R and B. Here, some skin pixels that fail to satisfy the criteria are likely to be left undetected. Moreover, some non-skin pixels remain in the combined regions.

The abovementioned scenario leads us to incorporate more criteria to accurately detect skin pixels. To gather enough evidence we vary the criteria to generate different SoIs. The criteria are established using different mathematical expressions consisting of R, G and B. Observing the distributions of skin and non-skin pixel colors for each of these criteria, we select as SoIs only those criteria that exhibit a good difference between the two distributions. For instance, we plot skin and non-skin counts with respect to the $|R-G|$ values. This plot is presented in Fig. 5(a), which leads to the SoI of $28 < |R-G| < 130$. Fig. 5(b) shows the region covered by this criterion.

Similarly, we find the SoI $45 < |R-B| < 187$ as shown in Fig. 6.

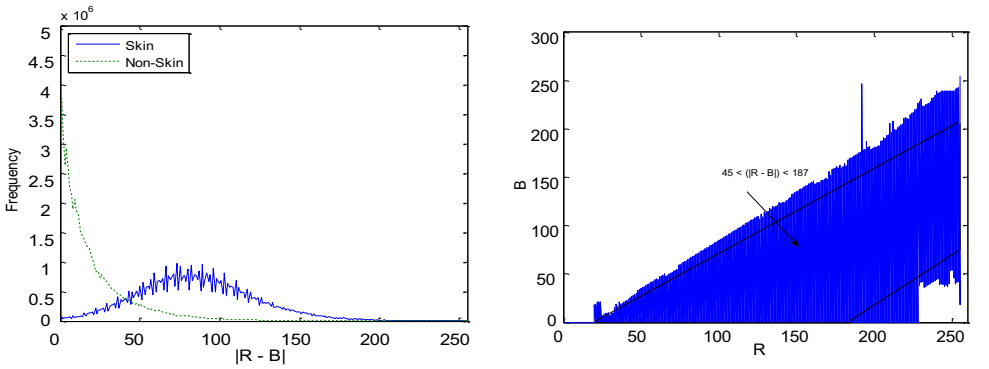
The value of R is usually larger than G and B in the skin region. Thus, we take ($R > G$ and $R > B$) as another SoI. Other criteria (such as $|G-B|$) do not exhibit much discrimination between skin and non-skin pixels, and hence are not chosen.

Finally, we take the six discriminating criteria as SoIs in our proposed DSSD. The first column of Table 3 presents the selected SoIs obtained using the training set presented in Section 6.1. Note that such measures are not new and have been used



(a) Distribution of skin and non-skin colors (b) Coverage of the criteria in RG space

Fig. 5. Clustering based on $|R-G|$.



(a) Distribution of skin and non-skin colors (b) Coverage of the criteria in RB space

Fig. 6. Clustering based on $|R-B|$.

Table 3
Mass values of the selected SoIs.

SoI	$m_i(\{\text{skin}\})$	$m_i(\{\text{non-skin}\})$
$R > 140$	0.364339536	0.524533342
$75 < G < 193$	0.147315460	0.430099211
$43 < B < 161$	0.222264545	0.440255872
$28 < \|R-G\| < 130$	0.546243640	0.675261991
$45 < \|R-B\| < 187$	0.496808781	0.673853774
$R > G$ and $R > B$	0.387888000	0.968090001

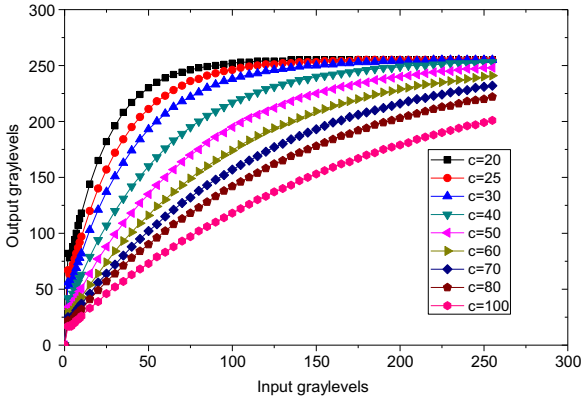


Fig. 7. The effect of c in the projected gray levels.

in several articles [1,4,18]. However, our procedure for selecting the thresholds is different from that used in previous studies. While most existing approaches set the thresholds so that they can encompass the skin cluster irrespective of the non-skin distributions, our thresholds are set at the point where a discrimination between skin and non-skin pixel distribution exists between the two sides of the threshold value. Thus, we use the criteria such that each threshold can classify both skin (when the criterion is satisfied) and non-skin colors (when the criterion is not satisfied). Note that we have also proposed a generalized DSSD that does not require any threshold selection, described in detail in Section 5.3.

As mentioned in Section 4.1, using a large number of images for the selection of the SoIs is expected to allow the modeling of a variety of lighting conditions. However, a large change in illumination requires some preprocessing to adjust the intensities for better detection. Several nonlinear preprocessing methods [46,86–88] that make use of different color spaces such as HSV (Hue, Saturation and Value) and LUX (Logarithmic hUe eXtension) may be applicable in this regard. In this paper, intensities are adjusted by adjusting the brightness component of the HSV color space using the nonlinear transformation function stated in Eq. (6). After this adjustment, the color is converted back to the RGB color space for use in different steps of the proposed schemes.

$$I = \frac{2}{1 + e^{-2v/\eta}} - 1, \tag{6}$$

where I is the adjusted illumination, v is the V component of the HSV color space, $\eta = c\log(v)$ controls the curvature of the sigmoid function based on a constant c, and the value of c is determined based on the image contents. Finally, I is converted to the grayscale range [0,255].

Fig. 7 demonstrates that a small value of c projects the very dark pixels to much brighter gray levels, while a large c moves the brightest pixels to lower gray levels as expected from an illumination correction step. Under normal illumination conditions, the gray levels (V component) of the pixels are usually at intermediate levels where the skin detection also performs well. However, under extremely low or high illumination conditions, the pixels become too dark or too bright, respectively. Therefore, a small value of c will help to detect dark pixels and a higher c value will support the detection of very bright pixels. Our transformation function is similar to the method described in [46], but is simpler and performs well as a preprocessing step of skin detection for images with illumination problems.

5. Dempster–Shafer based skin detection (DSSD)

DSSD comprises three basic steps — selection of SoIs, determination of mass values (m) provided by the selected SoIs, and combination of the mass values to make the final decision. The SoI selection procedure was described earlier in Section 4.2. Here we discuss how to calculate mass values and make a decision by combining the mass values using DST. We then propose a generalized version of DSSD where SoIs do not require threshold selection.

5.1. Determination of mass values

The mass function defines the values representing the confidences of the SoIs in classifying pixels. To statistically compute the mass values (m) of the selected SoIs we use the four well-known notations *TP*, *TN*, *FP* and *FN*, where *TP* and *TN* represent the total number of pixels correctly classified whereas *FP* and *FN* are the number of pixels incorrectly classified as skin and non-skin pixels, respectively.

We measure *TP* and *FP* for each of the criteria defined in Section 4.2. We then calculate the *True Positive Rate* (*TPR*) and *False Positive Rate* (*FPR*) according to Eqs. (7) and (8), which represent the accuracy and inaccuracy in classification, respectively:

$$TPR = \frac{TP}{TP + FP}, \tag{7}$$

$$FPR = \frac{FP}{TP + FP}. \tag{8}$$

Finally, we determine the *Absolute Detection Rate for skin* (ADR^s) following

$$ADR^s = TPR - FPR. \tag{9}$$

Here, the *TPR* represents the certainty of classifying a skin pixel in the skin region defined by an SoI. However, the certainty is reduced due to the presence of non-skin pixel colors in the skin region, and *FPR* measures this reduction. Hence, the difference between these two values indicates the net gain in certainty (a similar definition of the net gain in certainty can also be found in [44]), which is represented as ADR^s in Eq. (9). The value of ADR^s varies between 0 and 1, reaching a maximum when $FPR = 0$ and a minimum when $FPR = TPR$ (according to our method of selecting SoI boundaries, *FPR* can never be higher than *TPR* in skin regions). Thus, ADR^s can be regarded as a piece of evidence to be used as a mass value, as it captures the discrimination power of an SoI when making positive (skin) decisions. However, analyzing the distributions of skin and non-skin colors, one of our key findings is that the potential of the SoI is different within and outside the skin cluster. Hence, it is not reasonable to use the same mass value for either classification done by an SoI. For this reason, we define ADR^{ns}, the *Absolute Detection Rate for non-skin*, to quantify the performance of each criterion when it detects non-skin pixels.

We represent the absolute detection rate of the ith SoI by ADR_i^s or ADR_i^{ns} when it detects a skin or a non-skin color, respectively. This is done according to Eqs. (10) and (11), which follow Eqs. (7)–(9):

$$ADR_i^s = \frac{TP_i - FP_i}{TP_i + FP_i}, \tag{10}$$

$$ADR_i^{ns} = \frac{TN_i - FN_i}{TN_i + FN_i}. \tag{11}$$

Some statistical indices (for example, the Predictive Summary Index (PSI) [44], the Youden Index [47], etc.) are available in the literature to measure the relevance of tests. Such measures quantify the accuracies of both the positive and negative decisions made by a test and then calculate the overall performance of the test. Practical PSI values range from 0 to 1, where 0 reflects a useless test and 1 indicates a perfect test [84]. Appendix B demonstrates the relationship between the PSI and our measures of evidence, namely ADR_i^s and ADR_i^{ns}. Thus, ADR_i^s and ADR_i^{ns} provide good expressive measures of the correctness of a criterion's positive and negative decisions, respectively, in our work.

In our skin detection process, the frame of discernment ($\boldsymbol{\Theta}$) is {skin, non-skin} for each pixel. Hence, the power set contains $4(=2^2)$ subsets, which includes $\emptyset$, {skin}, {non-skin} and {skin, non-skin}. According to the basic principle of DST, $m(\emptyset) = 0$. For other subsets, we assign the mass values for each SoI i according to Algorithm 1, satisfying the fundamental criteria (Eq. (1)) of mass functions presented in Section 3.

Algorithm 1. Assignment of mass values

1. **if** a criterion i holds **then**
2. $\quad m_i(\{skin\}) = ADR_i^s$
3. $\quad m_i(\{nonskin\}) = 0$
4. $\quad m_i(\{skin, nonskin\}) = 1 - ADR_i^s$
5. **else**
6. $\quad m_i(\{skin\}) = 0$
7. $\quad (m_i\{nonskin\}) = ADR_i^{ns}$
8. $\quad m_i(\{skin, nonskin\}) = 1 - ADR_i^{ns}$
9. **end if**

Table 3 presents a set of mass values along with the selected SoIs, and Example 4 illustrates the usage of these values.

Example 4. If the condition $R > 140$ is true, then the mass values according to Algorithm 1 are $m_{R>140}(\{\text{skin}\}) = 0.36$, $m_{R>140}(\{\text{non-skin}\}) = 0.0$ and $m_{R>140}(\{\text{skin,non-skin}\}) = 0.64$ (for simplicity, only two-digits after the decimal point are shown). Conversely, if the condition is not true, then the mass values are $m_{R>140}(\{\text{non-skin}\}) = 0.52$, $m_{R>140}(\{\text{skin}\}) = 0.0$, and $m_{R>140}(\{\text{skin,non-skin}\}) = 0.48$.

5.2. Combining the outcomes of the SoIs using DST

To obtain a reliable decision (i.e., even if a criterion misclassifies a pixel, the combination of all other decisions may overcome the possible error), we make use of six different classification criteria (SoIs) as described in Section 4.2, along with their mass values as presented in Section 5.1. As stated in DST, Eq. (4) in Section 3 combines the mass values of the individual SoIs. Example 5 (when all SoIs agree) and Example 6 (when disagreement exists) show two simple illustrations of such a combination of evidence for two SoIs.

Example 5. Suppose we have two SoIs, $R > 140$ and $28 < |R - G| < 130$ and they are both satisfied for a given color. The mass values are set as $m_{R>140}(\{\text{skin}\}) = 0.36$; $m_{R>140}(\{\text{non- skin}\}) = 0.0$; $m_{R>140}(\Theta) = 0.64$; $m_{28<|R-G|<130}(\{\text{skin}\}) = 0.55$; $m_{28<|R-G|<130}(\{\text{non-skin}\}) = 0.0$; $m_{28<|R-G|<130}(\Theta) = 0.45$. Table 4 shows these mass values along with the pair-wise combinations of different hypotheses.

According to Eq. (4), the degree of conflict $K = 0$. Hence, the normalization factor $1 - K = 1 - 0 = 1$. The combined mass values for different hypotheses are:

$$m(\{\text{skin}\}) = 0.198 + 0.162 + 0.352 = 0.712,$$
$$m(\{\text{non-skin}\}) = 0.0,$$
$$m(\Theta) = 0.288.$$

Therefore, based on Eq. (5) in Section 3, the given color is declared as skin.

Example 6. Consider that $R > 140$ is true but $28 < |R - G| < 130$ is false for a given color. The mass values are set as $m_{R>140}(\{\text{skin}\}) = 0.36$; $m_{R>140}(\{\text{non-skin}\}) = 0.0$; $m_{R>140}(\Theta) = 0.64$; $m_{28<|R-G|<130}(\{\text{skin}\}) = 0.0$; $m_{28<|R-G|<130}(\{\text{non-skin}\}) = 0.67$; $m_{28<|R-G|<130}(\Theta) = 0.33$. Table 5 shows these mass values along with the pair-wise combinations of different hypotheses.

According to Eq. (4) in Section 3, the combined mass values for different hypotheses (without normalization) are—

$$m(\{\text{skin}\}) = 0.118,$$
$$m(\{\text{non-skin}\}) = 0.428,$$
$$m(\Theta) = 0.211.$$

Here, the degree of conflict $K = 0.241$. Hence, after normalization we get:

$$m(\{\text{skin}\}) = 0.155,$$
$$m(\{\text{non-skin}\}) = 0.564,$$
$$m(\Theta) = 0.281.$$

Dempster's combination rule thus takes the total conflicting mass and then equally redistributes it to all focal elements. Finally, based on Eq. (5), the given color is declared as non-skin.

Table 4
An example of combining mass values for Example 5.

		$m_{28<\|R-G\|<130}(\{\text{skin}\})$		$m_{28<\|R-G\|<130}(\{\text{non-skin}\})$		$m_{28<\|R-G\|<130}(\Theta)$	
		0.55		0.0		0.45	
$m_{R>140}(\{\text{skin}\})$	0.36	{skin}	0.198	$\emptyset$	0.00	{skin}	0.162
$m_{R>140}(\{\text{non-skin}\})$	0.0	$\emptyset$	0.000	{non-skin}	0.00	{non-skin}	0.000
$m_{R>140}(\Theta)$	0.64	{skin}	0.352	{non-skin}	0.00	Θ	0.288

Table 5
Pair-wise combinations of mass values for Example 6.

		$m_{28<\|R-G\|<130}(\{\text{skin}\})$		$m_{28<\|R-G\|<130}(\{\text{non-skin}\})$		$m_{28<\|R-G\|<130}(\Theta)$	
		0.0		0.67		0.33	
$m_{R>140}(\{\text{skin}\})$	0.36	{skin}	0.00	$\emptyset$	0.241	{skin}	0.118
$m_{R>140}(\{\text{non} - \text{skin}\})$	0.00	$\emptyset$	0.00	{non-skin}	0.000	{non-skin}	0.000
$m_{R>140}(\Theta)$	0.64	{skin}	0.00	{non-skin}	0.428	Θ	0.211

Example 7. Consider a pixel color k where $R = 206$, G = 143 and B = 2 (we have chosen this color from the skin region of the image in Fig. 9). Using our training set described in Section 6.1, $P(k|Skin) = 1.5e{-}08$ and $P(k|NonSkin) = 7.53e{-}08$ according to Eq. (A.1) in Appendix A. The ratio of these values is 0.1996, which cannot pass the test in Eq. (A.2) of the Bayesian classifier even if we set η as small as 0.2. Therefore, this color is considered as **non-skin**. On the other hand, for DSSD, two SoIs classify k as non-skin while all the rest classify it as skin. Using the mass values and the SoIs from Table 3, we get $m_{R>140}(\{\text{skin}\}) = 0.36$, $m_{75<G<193}(\{\text{skin}\}) = 0.14$, $m_{43>B>161}(\{\text{non-skin}\}) = 0.44$, $m_{28<|R-G|B<130}(\{\text{skin}\}) = 0.54$, $m_{45<|R-B|<187}(\{\text{non-skin}\}) = 0.67$, $m_{R>G \text{ and } R>B}(\{\text{skin}\}) = 0.38$. The mass values for other hypotheses can be calculated according to Algorithm 1.

From these values, we get $K = 0.00337$. Thus, using Eq. (4) we get $m(\{\text{skin}\}) = 0.1556$, $m(\{\text{non-skin}\}) = 0.12348$ and $m(\boldsymbol{\Theta}) = 0.02758$. Hence, using Eq. (5) this pixel color is detected as **skin**.

Example 7 shows the limitations of the Bayesian classifier and the superiority of the proposed DSSD method. The Bayesian classifier depends fully on the conditional probability of the particular color according to the training set. Hence, it cannot cope with missing information and/or ambiguity. On the other hand, DSSD can eliminate such a problem thanks to the combination rule of DST that effectively combines several SoIs to make a better decision. Moreover, depending on the decision-making threshold, the Bayesian classifier may lead to different decisions for the same color. However, the DSSD will always lead to the same decision for a given color.

5.3. Generalization of DSSD

In DSSD, an SoI divides the skin and non-skin distributions into two parts and calculates their masses. A pixel color is checked against the criterion, and the mass values corresponding to the part it falls into are assigned according to Algorithm 1. The use of such an SoI has two shortcomings. First, it incorporates threshold(s) that must be set. Second, an SoI in DSSD calculates only one mass value for a partition. In other words, all the colors falling into the same partition are assigned the same value irrespective of the position and distribution. For example, consider that w, x, y and z are four different G-values in Fig. 8. The SoI ($75 < G < 193$) will declare both x and y as skin with the same mass. However, the distributions clearly show that the confidence of such a decision is much greater for x than for y. A similar comment can also be made for w and z.

To eliminate such shortcomings we extend DSSD to a generalized DSSD (GDSSD) where the assignment of mass values is more reliable. The SoIs in the GDSSD approach are similar to those in DSSD with the exception that no threshold is used for the SoIs in the case of GDSSD. The mass is calculated for every value in the horizontal axis in the distributions.

Suppose that the total counts of skin and non-skin pixels are ns and nns, respectively, for the value x in the distributions illustrated in Fig. 8. Now, instead of using the criterion $75 < G < 193$ as in DSSD, we employ $ns > nns$ as a classifier. When the criterion $ns > nns$ holds (or does not hold), the color is classified by the SoI as skin (or non-skin), and the mass value is calculated according to Eq. (10) (or Eq. (11)) in Section 5.1 using ns and nns as TP and FP (or, FN and TN), respectively. The rest of the steps in GDSSD are the same as in DSSD.

This approach means that the GDSSD does not require the selection of boundaries of SoIs. Moreover, another benefit is that now we have a more accurate mass selection method for every color that helps to reduce uncertainty better than in the DSSD method. Example 8 illustrates the performance gain obtained from GDSSD over DSSD and the Bayesian method.

Example 8. Consider a skin pixel color c where $R = 254$, G = 229 and B = 163. In the case of a Bayesian classifier, $P(c|Skin) = 0$ because there was no skin sample for this particular color in the training dataset. Thus, the likelihood ratio becomes 0.0 and cannot pass the test in Eq. (A.2) in Appendix A. Therefore, this pixel color is considered as **non-skin**.

For DSSD, three SoIs classify this pixel as skin and the three remaining ones classify it as non-skin. Using the mass values and the SoIs from Table 3, we get $m_{R>140}(\{\text{skin}\}) = 0.36$, $m_{75<G<193}(\{\text{non-skin}\}) = 0.43$, $m_{43>B>161}(\{\text{non-skin}\}) = 0.44$, $m_{28<|R-G|<130}(\{\text{non-skin}\}) = 0.67$, $m_{45<|R-B|<187}(\{\text{skin}\}) = 0.49$, $m_{R>G \text{ and } R>B}(\{\text{skin}\}) = 0.38$. The respective mass values for other hypotheses can be calculated according to Algorithm 1.

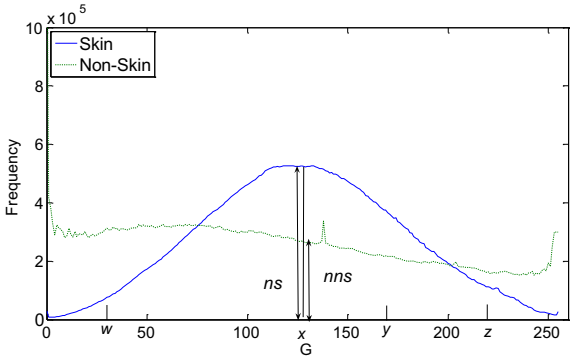


Fig. 8. Distribution of skin and non-skin clusters in G space.

From these values, we get $K = 0.00898$. Thus, using Eq. (4) we get $m(\{\text{skin}\}) = 0.0841$, $m(\{\text{non-skin}\}) = 0.1771$ and $m(\Theta) = 0.0204$.

Hence, using Eq. (5) the given color is detected as **non-skin**.

For GDSSD, we get $m_R(\{\text{skin}\}) = 0.05$, $m_G(\{\text{non-skin}\}) = 0.32$, $m_B(\{\text{skin}\}) = 0.02$, $m_{|R-G|}(\{\text{non-skin}\}) = 0.25$, $m_{|R-B|}(\{\text{skin}\}) = 0.63$, $m_{R>G \text{ and } R>B}(\{\text{skin}\}) = 0.38$, and from these values, we get $K = 0.00002$. Thus, using Eq. (4) we get $m(\{\text{skin}\}) = 0.4041$, $m(\{\text{non-skin}\}) = 0.1025$ and $m(\Theta) = 0.1073$.

Hence, using Eq. (5) this pixel color is detected as **skin**.

6. Comparative analysis

This section first presents the dataset followed by the evaluation criteria that we have used. We then compare the performances of our proposed methods with two state of the art skin detection methods, namely Mixture of Gaussian (MoG) and Bayesian skin classifiers. Both of these methods have gained a great deal of attention in the research community due to their acceptable detection rates. Until today, the performance of the Bayesian-based method presented in [2] remains highly competitive as compared to the other available methods [75,78]. We have also included the comparative performances of an explicit threshold-based method, RGB_ETM, [4,6] and a well known Dempster–Shafer-based classifier that uses a neural network (NNDST) [66] in our results section (to the best of our knowledge, we are the first to introduce NNDST[3] for skin detection).

6.1. Source of the experimental data

A large amount of data is needed for experiments based on statistics. The Compaq Skin and Non-skin Database [2] includes a large amount of skin and non-skin images. The database contains more than 14,000 images consisting of almost two billion pixels. The database also includes manually labeled skin and non-skin masks for all images that contain skin regions. These masks serve as ground truths in evaluating the detected skin regions. We have also used the IBTD [22] database containing 554 images that include 21,076,098 skin pixels, the University of Chile skin database with 103 images containing 5,744,330 skin pixels [31] and the Caltech Frontal Face Dataset [45] for testing the methods' performance under various conditions.

Our training set includes 4650 images from the Compaq database — 3050 images containing 51,744,044 skin pixels in total and 1600 images consisting of 189,195,041 non-skin pixels. We have used a large number of non-skin pixels to capture a good distribution of the non-skin colors in nature. However, the greater number of non-skin pixels in the training set might lead them to dominate over the skin pixels when calculating the mass values. Thus, the total number of non-skin pixels was normalized to that of skin pixels. This normalization helps to avoid the discrimination, retaining the shape of the distribution of non-skin pixels. The test set, also picked from the Compaq database, consists of 3000 images containing 26,787,957 skin and 250,989,469 non-skin pixels, and these test images do not have any image in common with the training images. The aforementioned train and test dataset is used to generate the results in this paper unless otherwise specified.

6.2. Evaluation criteria

To evaluate the strength of the proposed method and to compare its performance with other well-established proposals, we have calculated three different criteria as presented in [48]: Correct Detection Rate (CDR) — percentage of skin pixels correctly classified, False Detection Rate (FDR) — percentage of non-skin pixels incorrectly classified as skin pixels, and Classification Rate (CR) — percentage of pixels correctly classified as a whole.

Besides these we have compared the methods using the well-accepted F-measure (F) [79], defined as

$$F = 2 \cdot \frac{precision * recall}{precision + recall}, \tag{12}$$

where *recall* is same as CDR and *precision* is defined as

$$precision = \frac{TP}{TP + FP}. \tag{13}$$

6.3. Results and discussion

For each experiment we first used the training samples to select the SoIs (for DSSD) along with their boundaries as described in Section 4.2, and then computed the masses (m). We have also computed the mass values for GDSSD and trained the other methods with the same training set. The decision-making thresholds for Bayesian as well as MoG classifiers are set at the levels that produce the highest accuracy.

[3] The source code of the original NNDST is available at http://www.hds.utc.fr/~tdenoeux/dokuwiki/doku.php?id=en:software.

Table 6
Performance (%) of different skin detection methods.

Method	Precision	CDR (Recall)	FDR	CR
Bayesian classifier	82.15	89.48	19.43	81.43
MoG classifier	81.87	89.10	19.73	81.12
RGB_ETM	80.23	87.01	21.88	79.15
NNDST	82.53	90.01	19.07	81.80
DSSD	83.77	91.56	17.73	83.32
GDSSD	84.95	91.96	16.28	84.51

We have experimentally found that the proposed methods are able to determine a clearer skin segment in an image irrespective of ethnicity, background and illumination conditions with higher detection rates and lower false detection rates than the other methods. Table 6 summarizes the outcomes of the different methods.

From Table 6, it can be observed that the proposed methods (both DSSD and GDSSD) show better performances with higher correct detection rates and normalized precision (the normalization procedure is described later) and lower false detection rates than the other methods. In GDSSD the masses are assigned more meaningfully than in DSSD, which leads to better performance. Again, we have performed the skin detection test following the Bayesian fusion method described in [91] where we have combined the information (conditional probabilities) provided by the same six SoIs. The final decision is taken in the favor of the class (skin or non-skin) having the higher probability. This method has resulted in a decreased CR (78.47%) where the CDR and FDR are 91.18% and 24.37%, respectively, for the same dataset used to generate the Table 6. This implies that the proposed belief function and the uncertainty management capability of DST have a great impact on the obtained results of DSSD and GDSSD methods.

Table 7 shows the *F*-measures (F) of the methods. The right-most column (NF) shows the *F*-measures when the total number of non-skin pixels in the test set is normalized to that of skin pixels (we have done this normalization because the large number of non-skin pixels dominates the skin pixels in the test set). The *F*-measures also demonstrate the superiority of the proposed methods.

To obtain more robust and reliable results, we have created a large set of images combining our train and test sets from the Compaq database and the images from the IBTD and the University of Chile skin databases. We have then run a 5-fold cross validation test on this set, with the results presented in Table 8. Here we also find that the proposed DSSD and GDSSD methods exhibit much better performance than the other methods. The performance of NNDST is comparable with the performance of the proposed methods (especially with DSSD) in some cases. The main reason for the success of the NNDST method is that it also works under the DST formalism, and thus has ability to handle uncertainty and incompleteness in the training data. However, the major problem with this method is the long training and classification times as mentioned earlier. Hence, the proposed methods are much preferred since they achieve better accuracies with lower training and classification times.

Figs. 9–12 present a few illustrative images from the test dataset along with the outcomes of the different methods. These figures also demonstrate that our proposed method can handle a variety of conditions better than the existing methods. Now we focus on several noteworthy observations of our experiments. In all the figures described next, the black pixels represent areas determined as non-skin. The key findings are summarized and discussed as follows:

6.3.1. Segmentation

For almost all the images, the proposed approaches yield very good outcomes by providing solid segments without many holes. The Bayesian classifier also detects most of the skin pixels successfully but leaves some holes/undetected pixels and

Table 7
F-measure (F) and normalized *F*-measure (NF) for different methods.

Method	TP	FP	TN	FN	F	NF
Bayesian	23969085	48777955	202211514	2818872	0.481	0.856
MoG	23867004	49519391	201470078	2920953	0.476	0.853
RGB_ETM	23332443	53864356	197125113	3455514	0.448	0.835
NNDST	24134132	47865334	203124135	2653825	0.488	0.861
DSSD	24527082	44509784	206913380	2260875	0.511	0.874
GDSSD	24634056	40865236	210124233	2153901	0.534	0.883

Table 8
Results of the 5-fold cross validation test.

Method	CDR	FDR	CR
Bayesian classifier	87.80	19.05	82.26
MoG classifier	87.32	19.39	81.90
RGB_ETM	85.42	21.78	79.61
NNDST	89.86	18.65	82.99
DSSD	90.17	18.03	83.55
GDSSD	91.12	16.10	85.29

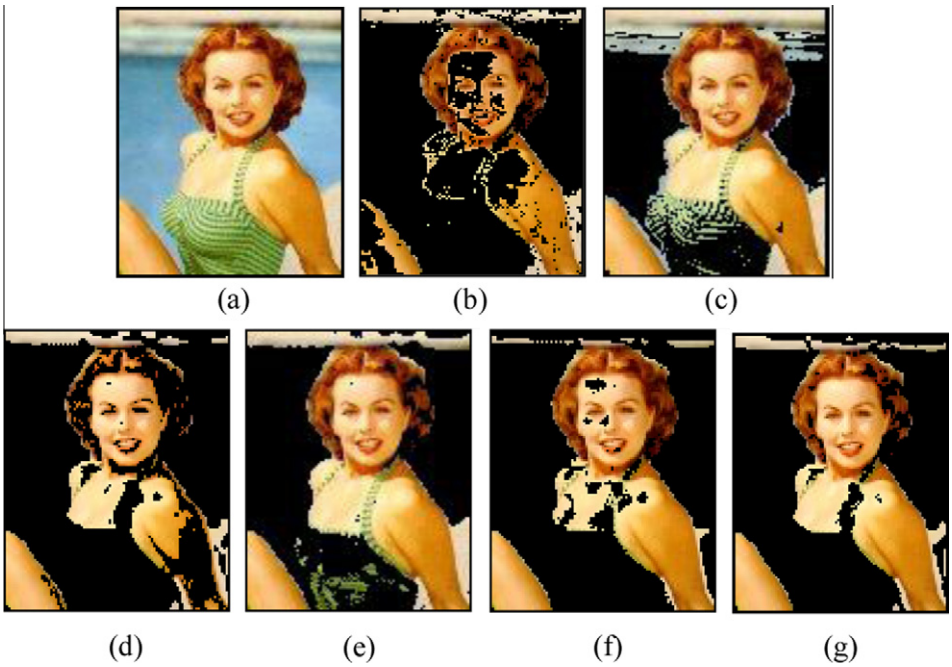


Fig. 9. Outcomes of the skin segmentation. (a) Original image (from the Compaq database), detections by (b) the Bayesian classifier, (c) the MoG classifier, (d) the RGB_ETM, (e) the NNDST, (f) the DSSD, and (g) the GDSSD.

includes some noisy pixels. The MoG method achieves a high detection rate but also identifies a number of non-skin pixels. The image in Fig. 9 is also presented in [2] demonstrating that the Bayesian classifier cannot detect the forehead and some other skin regions. The RGB_ETM method misses a good portion of the arm because the failure of a (or some) criterion to detect a skin pixel (for example, the condition $B > 20$ is not satisfied in many pixels in the missed region) is propagated to the output. However, the proposed schemes clearly overcome these problems by combining different evidences supplied by the selected criteria. Because the final decisions are made based on a weighted accumulation of individual decisions, the risk of erroneous decisions decreases and a good result is provided.

6.3.2. Complex background

Fig. 10 shows the results for images with complex backgrounds. The main difficulty here is to discard the skin-like non-skin pixels (and detect the genuine skin pixels). In these images, the DST-based methods perform very well. For example, the *F*-measures of the methods when applied to the left-most image in Fig. 10 are 0.52, 0.38, 0.48, 0.53, 0.61, and 0.64 from the top to the bottom image. The image in the second column contains skin regions with normal illumination as well as some very bright skin pixels that appear almost white. In this case, the Bayesian and MoG methods exhibit poor performance. However, the other four methods achieve almost similar overall performances (the *F*-measure is approximately 0.76). Although the RGB_ETM method detected the very bright skin pixels in the neck region, it also detected much background pixels, leading to a lower overall accuracy. Moreover, the ability of the RGB_ETM method to detect such bright skin pixels may also lead to the inclusion of whitish non-skin pixels in many cases. For the third image, the DST-based methods clearly identify most of the skin regions excluding the skin-like frame in the background and achieve the best performance (the F-measures of NNDST, DSSD and GDSSD are 0.64, 0.79 and 0.81, respectively).

6.3.3. Reddish background

Fig. 11 shows the performances of the methods in the presence of red-like non-skin colors. Most of the existing classifiers fail to reject these non-skin pixels because the red color remains dominant for skin data due to the presence of hemoglobin. However, the proposed approaches are mostly able to exclude such noisy pixels. For the second image in Fig. 11, all methods fail to exclude the red-like background. Although the Bayesian-based method has discarded most of the non-skin pixels, it has also excluded major portions of the chin and neck regions. Both the DSSD and GDSSD methods perform better than the other methods in discarding red-like non-skin pixels while retaining skin pixels.

6.3.4. Illumination

In the image in Fig. 12(a) taken from the Caltech Frontal Face Dataset, the brightness of the face region is very low. We preprocess the image using Eq. (6), and use the same preprocessed image for the different methods. Fig. 12 demonstrates that the proposed methods detect skin regions more accurately than the other methods.

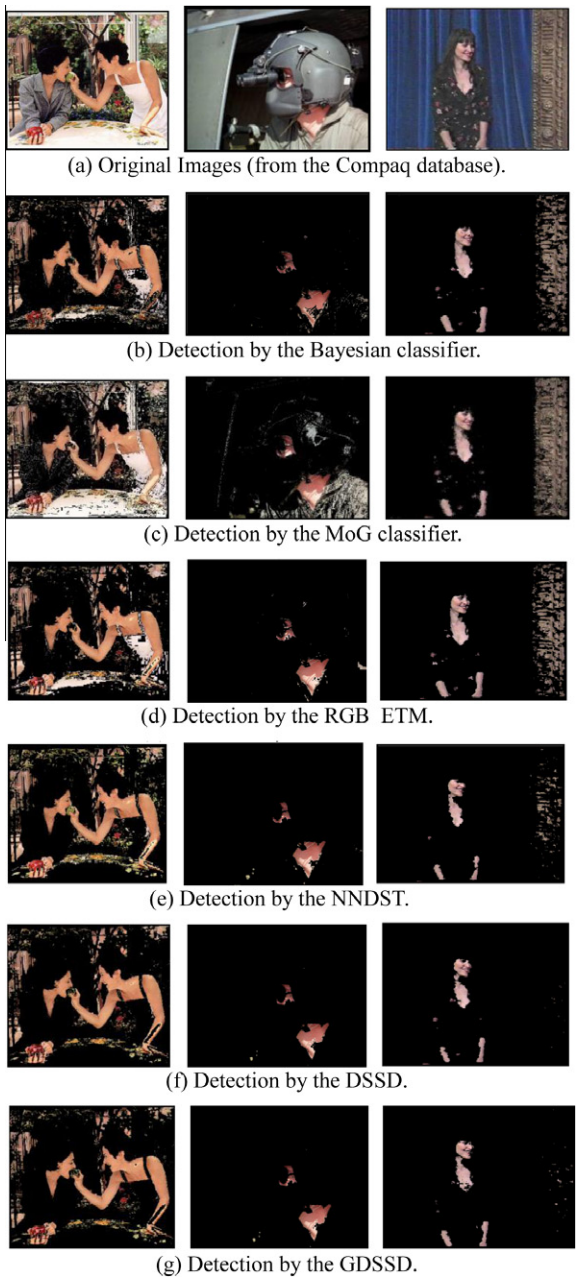


Fig. 10. Results of skin detection methods for a complex background.

6.3.5. Amount of training data

So far, we have discussed the performance of the methods when a large amount of training data is available. To investigate the behaviors of the methods when small amounts of data are used for training, we have randomly selected approximately 1% of the skin pixels and the same number of non-skin pixels from our training set. Fig. 13 presents typical plots of the skin and non-skin distributions for the selected pixels. These plots show that there are many intersecting points in the plots of skin and non-skin distributions because of the irregular shapes of the non-skin distribution, making it difficult to determine the boundaries for the SoIs in DSSDs. However, the skin samples show much better regularity despite very few samples.

To handle such situations, we propose two simple ways of using the DSSD method for small amounts of training data. The two alternatives differ only in the selection of the boundaries for the SoIs. For the first alternative, we put the boundary so that it includes all the skin samples within one standard deviation from each side of the mean of the skin distributions. Such boundaries include almost 70% of the skin samples (this selection method also results in similar boundaries to the SoIs

Fig. 11. Results of applying the skin detection methods to images with reddish non-skin objects.

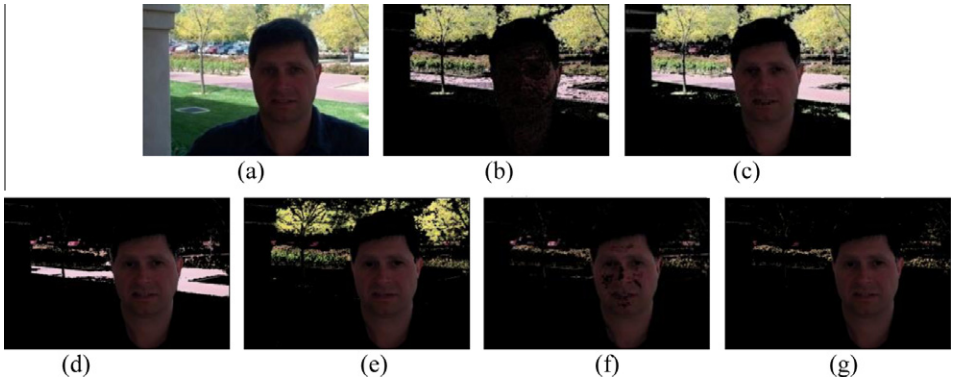
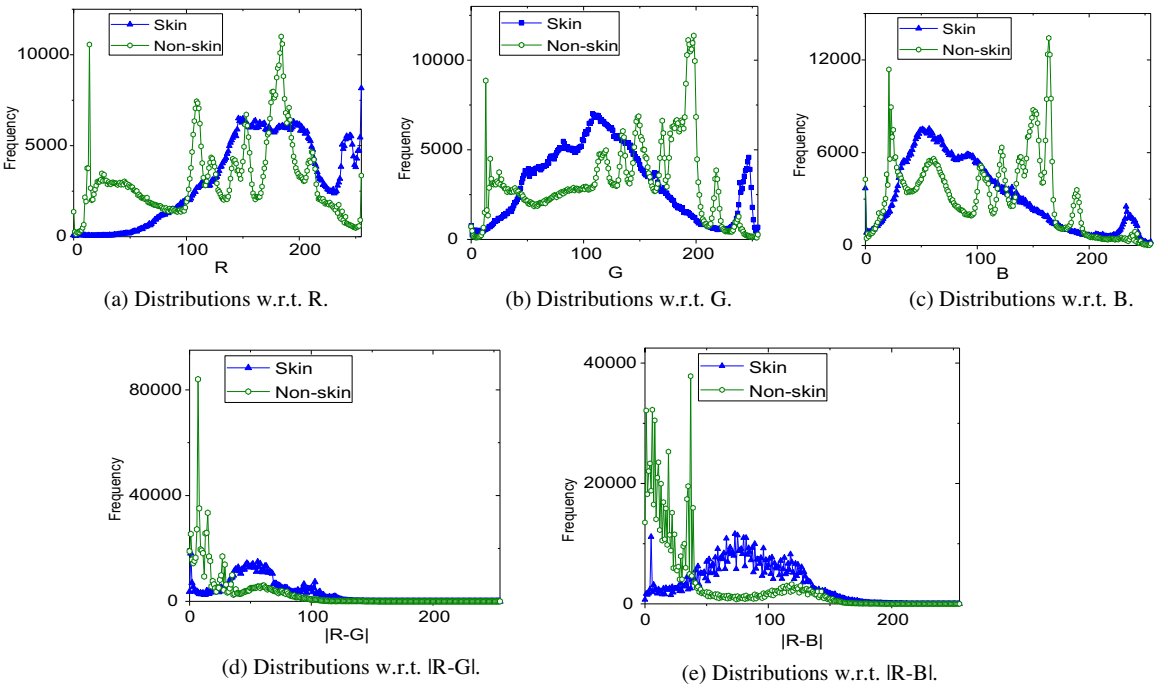


Fig. 12. Results of applying the skin detection methods to images with illumination problems. (a) Original image, detections by (b) the Bayesian classifier, (c) the MoG classifier, (d) the RGB_ETM, (e) the NNDST, (f) the DSSD, and (g) the GDSSD.

described in Section 4.2). For the second alternative, we keep the boundaries the same as in Section 4.2. After setting the boundaries, the rest of the process is the same as for DSSD. Table 9 presents the SoIs for the first alternative along with their boundaries and the mass values of both alternatives calculated using 1% of the data. The process of using GDSSD for a small amount of data is the same as described in Section 5.

We trained the methods using the randomly chosen 1% of the training data and then verified the methods using the test dataset mentioned in Section 6.1. We repeated this process five times to investigate the consistency of our results, and the average outcomes are presented in Table 10. The results demonstrate the robustness of the proposed methods whereas the Bayesian-based method yields much worse outcomes. Although the use of 16^3 bins gives better accuracy, the performance is still quite poor. The NNDST method also shows a much more reasonable accuracy compared to the Bayesian-based methods.

We have also tested the methods on the University of Chile skin database [31] using a similar training and testing process as done before. The averaged results are shown in Table 11, which also demonstrates that the proposed methods perform better than the others.



(a) Distributions w.r.t. R. (b) Distributions w.r.t. G. (c) Distributions w.r.t. B.

(d) Distributions w.r.t. |R-G|. (e) Distributions w.r.t. |R-B|.

Fig. 13. Plots of the skin and non-skin distributions for 1% of the training data.

Table 9
The redefined boundaries of the SoIs (for the first alternative) and the mass values of the two alternatives when using 1% of the training data.

SoIs (for the first alternative)	First alternative		Second alternative	
	$m_i(\{\text{skin}\})$	$m_i(\{\text{non-skin}\})$	$m_i(\{\text{skin}\})$	$m_i(\{\text{non-skin}\})$
$R > 99$	0.178123328	0.3284074864	0.124434354	0.5954218481
$42 < G < 181$	0.048259981	0.0868313786	0.152559516	0.3632760050
$25 < B < 150$	0.066580714	0.1355648535	0.118848672	0.2973934352
$28 < \lvert R - G\rvert < 83$	0.412317699	0.6012579511	0.474181966	0.3560897363
$35 < \lvert R - B\rvert < 136$	0.574493217	0.6356666317	0.472379029	0.5768617910
$R > G$ and $R > B$	0.456460261	0.8410131201	0.456460261	0.8410131201

Table 10
Performance of the methods trained on the small (1%) dataset.

Method		CDR	FDR	CR
Bayesian classifier (256^3 bin)		61.35	24.13	74.47
Bayesian classifier (16^3 bin)		68.81	25.05	74.36
NNDST		85.83	23.93	77.01
GDSSD		89.97	22.72	78.51
DSSD	First alternative	87.48	19.35	81.31
	Second alternative	83.75	22.94	77.71

Table 11
Performance of the methods trained on a small dataset (1%) selected from the University of Chile skin database.

Method		CDR	FDR	CR
Bayesian Classifier (256^3 bin)		71.99	21.06	77.73
Bayesian Classifier (16^3 bin)		75.38	20.15	79.07
NNDST		76.31	20.16	79.20
GDSSD		84.77	20.88	80.09
DSSD	First alternative	77.12	17.96	81.19
	Second alternative	84.26	21.25	79.71

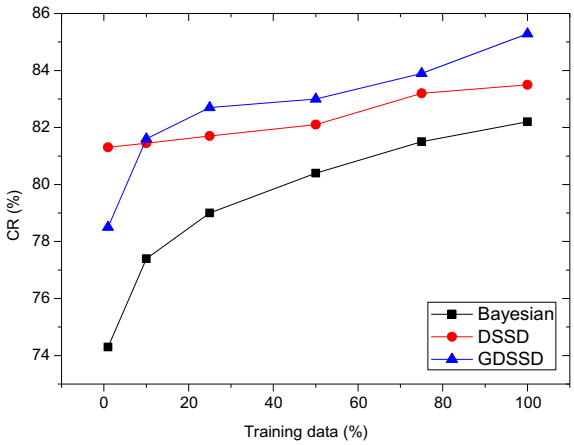


Fig. 14. The performances of different methods with different amounts of training data.

The abnormal fluctuations in the distributions demonstrated in Fig. 13 might change the mass values of the positive as well as negative decisions. We observe the effects of such changes (i.e., decreased TP and increased FP) in Table 10 in comparison with Table 6, where the same test set is used. Again, the effect of irregularities in non-skin distribution might be greater for GDSSD than for DSSD because DSSD calculates the mass values for the whole skin (and non-skin) region. Hence, even if some irregular behaviors are observed for non-skin, the effects become less as a whole, and thus we may expect reasonable mass values. On the other hand, GDSSD compares the skin and non-skin accumulations at every bin value of the SoIs. Therefore, the GDSSD is more likely to be affected by the instantaneous fluctuations than DSSD, and thus the overall accuracy of DSSD becomes greater for the small amount of training data.

The strength of the proposed methods is that they can still produce quite good results compared to the other methods for irregular distributions. Further, we can generate mass values of the SoIs with very few resources (training data and time) to handle any abnormal lighting situation.

Finally, we have tested the behavior of our proposed methods for different amounts of training data. For this, we varied the amount of training data (such as 1%, 10%, 25%, 50%, 75% and 100% of the training data) and used the same test set as described in Section 6.1. Up to 20% of the training data, we used the first alternative of DSSD and 16^3 bins for the Bayesian-based method. For greater amounts of training data we used the original DSSD and Bayesian-based methods. We also tested the performance of the GDSSD method on the same data using the original GDSSD as it requires no threshold or parameter tuning. Fig. 14 presents the outcomes of the different methods for different amounts of training data, with the results showing acceptable accuracies of the proposed methods compared to the state-of-the-art method.

7. Conclusion

This paper proposed a DST-based method named DSSD for detecting skin pixels in images. The DSSD method has successfully overcome the limitations of existing methods and also achieves quite reasonable accuracies with very small amounts of training data, which has many advantages for real life applications. However, DSSD is not fully threshold-free. The GDSSD method overcomes such threshold involvement and achieves the best performance as compared to the state-of-the-art methods.

Furthermore, the proposed method for calculating mass values may be applied along with the properly-selected SoIs with very little parameter tuning to solve different pattern classification problems under the DST combination rule to manage very high dimensional tasks as well as to handle large amounts of training data with very low computation and memory requirements.

Acknowledgements

We would like to express our deep gratitude to the anonymous reviewers for their constructive comments, with special thanks for inspiring us to pursue threshold-free SoIs, leading us to propose the generalized DSSD. We are also grateful to T. Denoeux for helping us in the implementation of NNDST.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0017151).

Appendix A

A.1. A brief overview of two well-known approaches

Most of the successful methods in the literature use different variants of Gaussian or Bayesian classifiers for skin detection. We discuss these two basic ideas here in brief.

A.1.1. The Bayesian (histogram) model

Jones and Rehg [2] first build two histograms for skin and non-skin from the pool of training pixels. Each histogram is then normalized by the total number of pixels in it according to Eq. (A.1):

$$\begin{aligned} P(Col|Skin) &= \frac{Count_{Skin}(Col),}{Total_{Skin}} \\ P(Col|Nonskin) &= \frac{Count_{Nonskin}(Col),}{Total_{Nonskin}} \end{aligned} \tag{A.1}$$

where Col is a vector representing a pixel's color (or the histogram bin where the color falls), and $P(Col|Skin)$ and $P(Col|Nonskin)$ represent the probabilities that the color Col belongs to the skin and non-skin classes, respectively. $Count_{skin}(Col)$ and $Count_{Nonskin}(Col)$ are the total number of pixels of the color Col in the skin and non-skin histograms, while $Total_{skin}$ and $Total_{Nonskin}$ represent the total number of pixels in the skin and non-skin training sets, respectively. Such a normalized histogram model represents the conditional probability, which is alternatively known as the likelihood.

Given the class conditional probabilities of skin and non-skin and using a standard likelihood ratio approach [14], a pixel is classified as skin if it is greater than some threshold η as described in Eq. (A.2).

$$\frac{P(Col|Skin)}{P(Col|NonSkin)} \geqslant \eta, \tag{A.2}$$

where η is a threshold that can be set as a tradeoff between true and false positives.

A.1.2. The Gaussian mixture model

Gaussian or Mixture-of-Gaussian (MoG) classifiers [11,15,24,29,30,50] construct an approximation of the probability distribution of the skin colors in the training set. MoG models are composed as the sum of some Gaussian kernels as presented in Eq. (A.3):

$$P(C) = \sum_{k=1}^{n} \alpha_k \frac{1}{(2\pi)^{\frac{3}{2}}|\mathrm{M}_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(C-m_k)^T \mathrm{M}_k^{-1}(C-m_k),} \tag{A.3}$$

where C is a color vector and the kth Gaussian is defined by the scalar weight α_k, mean vector m_k and diagonal covariance matrix M_k. The parameters are approximated by fitting the models to training data using expectation maximization (EM).

Different proposals have been made on how to make the final decision. The simplest method is to calculate the probability of each pixel's color in the image according to the trained skin model and if it is higher than the threshold it is declared as skin [83]. However, this approach does not consider the distribution of non-skin colors. Hence, in our work, we have followed a similar approach to that in Jones and Rehg [2]. We have trained two MoGs — one for skin and another for non-skin training samples. For a given color, the probabilities for both skin and non-skin samples are calculated first and then the decision is made by comparing the ratio of these probabilities with a threshold (similar to what is done in Eq. (A.2)).

Appendix B

ADR^s and ADR^{ns} are related to the Predictive Summary Index (PSI).

Proof. Let PPV and NPV denote Positive Predictive Value and Negative Predictive Value as

$$PPV = \frac{TP}{TP + FP}, \tag{B.1}$$

$$NPV = \frac{TN}{TN + FN}. \tag{B.2}$$

Linn et al. [44] proposed PSI as

$$\mathrm{PSI} = \mathrm{PPV} + \mathrm{NPV} - 1. \tag{B.3}$$

Using simple algebra, PSI can also be expressed (putting TP = a, FP = b, TN = d, FN = c) as

$$\begin{aligned} \mathrm{PSI} &= \frac{a}{a+b} + \frac{d}{c+d} - 1 = \frac{ad - bc}{(a+b)(c+d)} + 1 - 1 = \frac{1}{2} \cdot \frac{2ad - 2bc}{(a+b)(c+d)} = \frac{1}{2}\left(\frac{a-b}{a+b} + \frac{d-c}{d+c}\right) \\ &= \frac{1}{2}\left(\frac{\mathrm{TP} - \mathrm{FP}}{\mathrm{TP} + \mathrm{FP}} + \frac{\mathrm{TN} - \mathrm{FN}}{\mathrm{TN} + \mathrm{FN}}\right) = \frac{ADR^s + ADR^{ns}}{2}. \end{aligned}$$

Hence, PSI can also be expressed as an average of net gains in certainties in positive (ADR^s) and negative (ADR^{ns}) detections. □

References

[1] P. Kakumanu, S. Makrogiannis, N. Bourbakis, A survey of skin-color modeling and detection methods, Pattern Recognition 40 (3) (2007) 1106–1122.
[2] M. Jones, J. Rehg, Statistical color models with application to skin detection, IJCV 46 (1) (2002) 81–96.
[3] T. Denoeux, Analysis of evidence-theoretic decision rules for pattern classification, Pattern Recognition 30 (7) (1997) 095–1107.
[4] Francesca Gasparini, Silvia Corchs, Raimondo Schettini, Recall or precision-oriented strategies for binary classification of skin pixels, Journal of Electronic Imaging 17 (2) (2008).
[5] G. Shafer, A Mathematical Theory of Evidence, P.U. Press, Princeton, New Jersey, 1976.
[6] J. Kovac, P. Peer, F. Solina, 2D Versus 3D Colour Space Face Detection, in: 4th EURASIP Conference on Video/Image Processing and Multimedia Communications, Croatia, 2003, pp. 449–454.
[7] C. Garcia, G. Tziritas, Face detection using quantized skin colour regions merging and wavelet packet analysis, IEEE Transaction on Multimedia (1999) 264–277.
[8] I.-S. Hsieh, K.-C. Fan, C. Lin, A statistic approach to the detection of human faces in color nature scene, Pattern Recognition 35 (2002) 1583–1596.
[9] G. Gomez, E.F. Morales, Automatic feature construction and a simple rule induction algorithm for skin detection, in: Proceedings of the ICML Workshop on Machine Learning in Computer Vision, 2002, pp. 31–38.
[10] Y. Dai, Y. Nakano, Face-texture model based on SGLD and its application in face detection in a color scene, Pattern Recognition 29 (6) (1996) 1007–1017.
[11] S.L Phung, D. Chai, A. Bouzerdoum, A novel skin color model in YCbCr color space and its application to human face detection, in: Proceedings of IEEE International Conference on Image Proc. I, 2002, pp. 289–292.
[12] J.-H. Zheng, C.-Y. Hao, Y.-Y. Fan, X.-Y. Zang, Adaptive skin detection under unconstrained lighting conditions using a bigaussian model and illumination estimation, Image Analysis & Stereology 24 (2005) 21–33.
[13] M.-C. Chi, J.-A. Jhu, M.-J. Chen, H.263+ region-of-interest video coding with efficient skin color extraction, in: International Conference on Consumer Electronics (ICCE), 2006, pp. 381–382.
[14] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, 1972.
[15] J. Cai, A. Goshtasby, Detecting Human Faces in Colour Images, IVC 18 (1999) 63–75.
[16] J.L. Crowley, F.B. Crard, Multi-modal tracking of faces for video communications, in: IEEE Comput. Soc. Conf. CVPR, 1997, pp. 640–645.
[17] L.J. Liu, Y.H. Yang, Multiresolution color image segmentation, IEEE Trans. PAMI 16 (1994) 689–700.
[18] V. Vezhnevets, V. Sazonov, A. Andreeva, A survey on pixel-based skin color detection techniques, GRAPHICON (2003) 85–92.
[19] M. Soriano, B. Martinkauppi, S. Huovinen, M. Laaksonen, Skin detection in video under changing illumination conditions, in: International Conference on Pattern Recognition, vol. 1, 2000, pp. 839–842.
[20] D. Chai, K.N. Ngan, Face segmentation using skin colour map in videophone applications, IEEE Transactions on Circuits and Systems for Video Technology 9 (4) (1999) 551–564.
[21] J.W. Guan, D.A. Bell, Evidence Theory and its Applications, North-Holland, New York, 1991.
[22] Qiang Zhu, Ching-Tung Wu, Kwang-Ting Cheng, Yi-Leh Wu, An adaptive skin model and its application to objectionable image filtering, in: ACM International Conference on Multimedia, New York, USA, 2004, pp. 56–63.
[23] J. Yang, W. Lu, A. Waibel, Skin color modeling and adaptation, in: Proc. Asian Conf. Computer Vision II, 1998, pp. 687–694.
[24] S.J. McKenna, S. Gong, Y. Raja, Modelling facial colour and identity with Gaussian mixtures, Pattern Recognition 31 (12) (1998) 883–1892.
[25] J.Y. Lee, S.I. Yoo, An elliptical boundary model for skin color detection, in: Proceedings of the International Conference on Imaging Science, Systems, and Technology, 2002.
[26] R.L. Hsu, M. Abdel-Mottaleb, A.K. Jain, Face detection in color images, IEEE Trans. Pattern Anal. Machine Intell. 24 (5) (2002) 696–706.
[27] M.H. Yang, N. Ahuja, Detecting human faces in color images, in: ICIP98, 1998.
[28] J.C. Terrillon, M.N. Shirazi, H. Fukamachi, S. Akamatsu, Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, CFGR00 (2000) 54–61.
[29] N. Oliver, A. Pentland, F. Berard, Lafter: lips and face real time tracker, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR97, 1997, pp. 123–129.
[30] H. Greenspan, J. Goldberger, I. Eshet, Mixture model for facecolor modeling and segmentation, Pattern Recognition Letter 22 (14) (2001) 525–1536.
[31] J. Ruiz-del-Solar, R. Verschae, Robust skin segmentation using neighborhood information, in: Proceedings of International Confeence on Image Processing – ICIP, 2004, pp. 207–210.
[32] M.M. Fleck, D.A. Forsyth, C. Bregler, Finding naked people, Proceedings of European Conference on Computer Vision 2 (1996) 592–602.
[33] S. Marcel, S. Bengio, Improving face verification using skin color information, in: ICPR02, 2002.
[34] S. Srisuk, W. Kurutach, New robust face detection in color images, in: AFGR02, 2002, 291–296.
[35] B.D. Zarit, J.B. Super, F.K.H. Quek, Comparison of five color models in skin pixel classification, in: ICCV99, 1999.
[36] D. Chai, A. Bouzerdoum, A Bayesian approach to skin color classification in YCbCr color space, in: IEEE TENCON00, vol. 2, 2000, pp. 421–424.
[37] M.C. Shin, K.I. Chang, L.V. Tsap, Does colorspace transformation make any difference on skin detection? in: IEEE Workshop on Applications of Computer Vision, Orlando, FL, December, 2002, pp. 275–279.
[38] S.L. Phung, D. Chai, A. Bouzerdoum, A universal and robust human skin color model using neural networks, in: IJCNN01, 2001.
[39] H. Sahbi, N. Boujemaa, Coarse to fine face detection based on skin color adaptation, in: Workshop on Biometric Authentication, Lecture Notes in Computer Science, vol. 2359, 2002, pp. 112–120.
[40] M.J. Seow, D. Valaparla, V.K. Asari, Neural network-based skin color model for face detection, in: Proceedings of the 32nd Workshop on Applied Imagery Pattern Recognition, 2003.
[41] C. Chen, S.P. Chiang, Detection of human faces in colour images, IEE Proceedings of Vision Image Signal Process 144 (6) (1997) 384–388.
[42] D. Brown, I. Craw, J. Lewthwaite, A SOM based approach to skin detection with application in real time systems, in: BMVC, 2001.
[43] T.C. Lin, Partition belief median filer based on Dempster–Shafer theory for image processing, Pattern Recognition 41 (2008) 139–151.
[44] Shai Linn, Peter D Grunau, New patient-oriented summary measure of net total gain in certainty for dichotomous diagnostic tests, Epidemiologic Perspectives & Innovations 3 (2006) 11.
[45] M. Weber, Frontal Face Dataset, California Institute of Technology, 1999. <http://www.vision.caltech.edu/html-files/archive.html>.
[46] L. Tao, M.J. Seow, V.K. Asari, Nonlinear image enhancement to improve face detection in complex lighting environment, in: Proc. SPIE Electron. Imag., vol. 6064, 2006.
[47] E.J. Youden, Index for rating diagnostic tests, Cancer (1950) 32–35.
[48] S.L. Phung, A. Bouzerdoum, D. Chai, Skin segmentation using color pixel classification: analysis and comparison, IEEE Transaction on Pattern Recognition and Machine Intelligence 27 (1) (2005) 148–154.
[49] Jamal Ahmad Dargham, Ali Chekima, Sigeru Omatu, Chelsia Amy Doukim, Data fusion for skin detection, Artificial Life and Robotics 13 (2009) 438–441.
[50] M.-H. Yang, N. Ahuja, Gaussian mixture model for human skin colour and its applications in image and video databases, SPIE/EI&T Storage and Retrieval for Image and Video Databases (1999) 458–466.

[51] B. Li, X. Xue, J. Fan, A robust incremental learning framework for accurate skin region segmentation in color images, Pattern Recognition 40 (12) (2007) 3621–3632.
[52] Hung-Ming Sun, Skin detection for single images using dynamic skin color modeling, Pattern Recognition 43 (4) (2010) 1413–1420.
[53] Mohammad Shoyaib, M. Abdullah-Al-Wadud, Oksam Chae, A reliable skin detection using Dempster–Shafer theory of evidence, in: Lecture Notes in Computer Science, ICCSA, vol. 5593, 2009, pp. 764–779.
[54] C. Osswald, A. Martin, Discrete labels and rich foci in theory of evidence, in: International Conference on Information Fusion, 2008, pp. 1426–1423.
[55] R.J. Safranek, S. Gottschlich, A.C. Kak, Evidence accumulation using binary frames of discernment for verification vision, IEEE Transactions on Robotics and Automation 6 (4) (1990) 405–417.
[56] G. Shafer, J. Pearl, Readings in Uncertain Reasoning, Morgan Kaufmann Publishers Inc., San Francisco, CA, 1990.
[57] Z. Fu, J. Yang, W. Hu, T. Tan, Mixture clustering using multidimensional histograms for skin detection, ICPR04 (2004) 549–552.
[58] P. Smets, The combination of evidence in the transferable belief model, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 447–458.
[59] John Klein, Christèle Lecomte, Pierre Miché Hierarchical and conditional combination of belief functions induced by visual tracking, International Journal of Approximate Reasoning 51 (4) (2010) 410–428.
[60] Rafael Muñoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas, A. Carmona-Poyato, Multi-camera people tracking using evidential filters, International Journal of Approximate Reasoning 50 (5) (2009) 732–749.
[61] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut, Facial expression classification: an approach based on the fusion of facial deformations using the transferable belief model, International Journal of Approximate Reasoning 46 (3) (2007) 542–567.
[62] Isabelle Bloch, Defining belief functions using mathematical morphology – application to image fusion under imprecision, International Journal of Approximate Reasoning 48 (2) (2008) 437–465.
[63] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, International Journal of Approximate Reasoning 38 (2005) 133–147.
[64] Barry R. Cobb, Prakash P. Shenoy, On the plausibility transformation method for translating belief function models to probability models, International Journal of Approximate Reasoning 41 (3) (2006) 314–330.
[65] Isabelle Bloch, Some aspects of Dempster–Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account, Pattern Recognition Letters 17 (1996) 905–919.
[66] T. Denoeux, A neural network classifier based on Dempster–Shafer theory, IEEE Transactions on Systems, Man and Cybernetics A 30 (2) (2000) 131–150.
[67] L. Fouque, A. Appriou, W. Pieczynski, An evidential Markovian model for data fusion and unsupervised image classification, in: Proceedings of the Third International Conference on Information Fusion, 2000, pp. TUB4/25–TUB4/32.
[68] A.K. Brodzik, R.H. Enders, Semigroup structure of singleton Dempster–Shafer evidence accumulation, IEEE Transactions on Information Theory 55 (11) (2009) 5241–5250.
[69] B. Quost, M.-H. Masson, T. Denoeux, Classifier fusion in the Dempster–Shafer framework using optimized t-norm based combination rules, International Journal of Approximate Reasoning (2010), doi:10.1016/j.ijar.2010.11.008.
[70] F. Faux, F. Luthon, Robust face tracking using colour Dempster–Shafer fusion and particle filter, in: The 9th International Conference on Information Fusion, FUSION'06, 2006, pp. 1–7.
[71] T. Denoeux, A k-nearest neighbor classification rule based on Dempster–Shafer theory, IEEE Transactions on Systems, Man and Cybernetics 25 (05) (1995) 804–813.
[72] T. Denoeux, L.M. Zouhal, Handling possibilistic labels in pattern classification using evidential reasoning, Fuzzy Sets and Systems 122 (3) (2001) 47–62.
[73] T. Denoeux, P. Smets, Classification using belief functions: the relationship between the case-based and model-based approaches, IEEE Transactions on Systems, Man and Cybernetics B 36 (6) (2006).
[74] B. Quost, T. Denoeux, M.-H. Masson, Pairwise classifier combination using belief functions, Pattern Recognition Letters 28 (5) (2007) 644–653.
[75] Z. Xu, M. Zhu, Color-based skin detection: survey and evaluation, in: Proceedings of the 12th International Multi-Media Modelling Conference, Beijing, China, 2006, pp. 143–152.
[76] H. Larochelle, Y. Bengio, J. Louradour, P. Lamblin, Exploring strategies for training deep neural networks, Journal of Machine Learning Research 1 (2009) 1–40.
[77] Q. Zhu, K.-T. Cheng, C.-T. Wu, A unified adaptive approach to accurate skin detection, in: International Conference on Image Processing, ICIP '04, 2004, pp. 1189–1192.
[78] K. Nallaperumal, S. Ravi, C.N.K. Babu, R.K. Selvakumar, A.L. Fred, C. Seldev, S.S. Vinsley, Skin detection using color pixel classification with application to face detection: a comparative study, in: ICCIMA 2007, 2007, pp. 436–441.
[79] C.J. van Rijsbergen, Information Retrieval, second ed., Butterworth, London, 1979.
[80] A. Appriou, Uncertain data aggregation in classification and tracking processes, in: B. Bouchon-Meunier (Ed.), Aggregation and Fusion of Imperfect Information, Physica-Verlag, Heidelberg, Germany, 1998, pp. 231–260.
[81] Ph. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning 9 (1993) 1–35.
[82] A. Motro, P. Smets, Uncertainty Management in Information Systems: From Needs to Solutions, Kluwer Academic Publishers, 1997. pp. 225–254.
[83] T.S. Jebara, A. Pentland, Parametrized structure from motion for 3D adaptive feedback tracking of faces, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Puerto Rico, 1997, pp. 144–153.
[84] D. Porock, D. Parker-Oliver, G.F. Petroski, Marilyn Rantz, Research article The MDS Mortality Risk Index: the evolution of a method for predicting 6-month mortality in nursing home residents, BMC Research Notes 3 (2010) 200.
[85] N. Kumar, P.N. Belhumeur, S.K. Nayar, FaceTracer: a search engine for large collections of images with faces, in: European Conference on Computer Vision (ECCV), Marseille, France, 2008, pp. 340–353.
[86] F. Luthon, B. Beaumesnil, N. Dubois, LUX color transform for mosaic image rendering, in: IEEE International Conference on Automation Quality and Testing Robotics (AQTR), vol. 3, 2010, pp. 1–6.
[87] D. Alleysson, J. Hérault, Variability in color discrimination data explained by a generic model with nonlinear and adaptive processing, Color Research and Application 26 (s1) (2001) s225–s229.
[88] M. Liévin, F. Luthon, Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of face features in video, IEEE Transaction on Image Processing 13 (1) (2004) 63–70.
[89] J. Brand, J. Mason, A comparative assessment of three approaches to pixel level human skin-detection, ICPR01 1 (2000) 1056–1059.
[90] T.S. Caetano, S.D. Olabarriaga, D.A.C. Barone, Performance evaluation of single and multiple-Gaussian models for skin-color modeling, in: SIBGRAPI02, 2002
[91] I. Bloch, Information combination operators for data fusion: a comparative review with classification, IEEE Transactions on Systems, Man and Cybernetics, Part A 26 (1) (1996) 52–67.